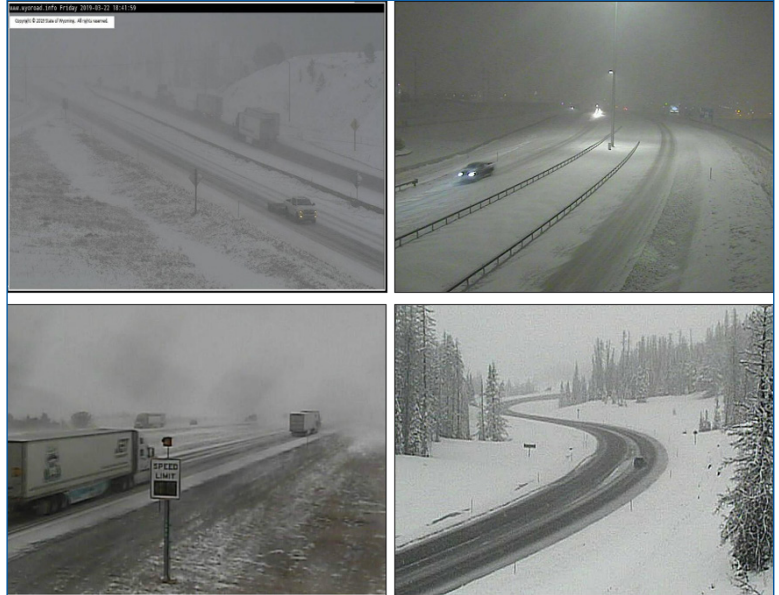


MOUNTAIN-PLAINS CONSORTIUM

MPC 22-485 | M. Ahmed, M. Khan and A. Das

AUTOMATED REAL-TIME
WEATHER DETECTION
SYSTEMS USING
ARTIFICIAL INTELLIGENCE



A University Transportation Center sponsored by the U.S. Department of Transportation serving the Mountain-Plains Region. Consortium members:

Colorado State University
North Dakota State University
South Dakota State University

University of Colorado Denver
University of Denver
University of Utah

Utah State University
University of Wyoming

Technical Report Documentation Page

1. Report No. MPC-630	2. Government Accession No.	3. Recipient's Catalog No.	
4. Title and Subtitle Automated Real-Time Weather Detection Systems Using Artificial Intelligence		5. Report Date July 2022	
		6. Performing Organization Code	
7. Author(s) Mohamed M. Ahmed, Ph.D., PE Md Nasim Khan, Ph.D., PE Anik Das, Ph.D., PE		8. Performing Organization Report No. MPC 22-485	
9. Performing Organization Name and Address Department of Civil and Architectural Engineering University of Wyoming 1000 E. University Ave Laramie, WY		10. Work Unit No. (TRAIS)	
		11. Contract or Grant No.	
12. Sponsoring Agency Name and Address Mountain-Plains Consortium North Dakota State University PO Box 6050, Fargo, ND 58108		13. Type of Report and Period Covered Final Report	
		14. Sponsoring Agency Code	
15. Supplementary Notes Supported by a grant from the US DOT, University Transportation Centers Program			
16. Abstract Adverse weather has long been recognized as one of the major causes of motor vehicle crashes due to its negative impact on visibility and road surface. Providing drivers with real-time weather information is therefore extremely important to ensure safe driving in adverse weather. However, identification of road weather and surface conditions is a challenging task because it requires the deployment of expensive weather stations and often needs manual identification and/or verification. Therefore, the primary objective of this research was to develop cost-effective systems capable of providing accurate weather and surface conditions in real time. First, a trajectory-level weather detection system was developed using only a single video camera mounted on the dashboard of the participant vehicles. Two texture-based features, histogram of oriented gradient (HOG) and local binary pattern (LBP), were extracted from images and used as classification parameters to train the weather detection models using several machine learning classifiers, such as gradient boosting (GB), random forest (RF), and support vector machine (SVM). In addition, a unique multilevel model, based on a hierarchical structure, was also proposed to increase detection accuracy. Evaluation results revealed that the multilevel model provided an overall accuracy of 89.2%, which is 3.2%, 7.5%, and 7.9% higher compared with the SVM, RF, and GB model, respectively, using the HOG features. Considering the LBP features, the multilevel model also produced the best performance with an overall accuracy of 91%, which is 1.6%, 8.6%, and 9% higher compared with the SVM, RF, and GB models, respectively. Afterward, the existing roadside webcams in Wyoming were utilized to develop several robust weather and surface detection systems by applying advanced deep learning techniques. Most U.S. departments of transportation (DOTs), including Wyoming DOT (WYDOT), have installed roadside webcams mostly for operational awareness. This study leveraged these easily accessible data sources to develop affordable automatic road weather and surface condition detection systems. The developed detection models are focused on three weather conditions: clear, light snow, and heavy snow; and three surface conditions: dry, snowy, wet/slushy. Several pre-trained convolutional neural network (CNN) models, including AlexNet, GoogLeNet, and ResNet18, were applied with proper modification via transfer learning to achieve the classification tasks. The best performance was achieved using ResNet18 architecture with an unprecedented overall detection accuracy of 97% for weather detection and 99% for surface condition detection. The proposed study has the potential to provide more accurate and consistent weather information in real time that can be made readily available to be used by road users and other transportation agencies. The proposed models could also be used to generate temporal and spatial variations of adverse weather for proper optimization of maintenance vehicle routes and time.			
17. Key Word artificial intelligence, detection and identification technologies, real time information, snow, weather conditions		18. Distribution Statement Public distribution	
19. Security Classif. (of this report) Unclassified	20. Security Classif. (of this page) Unclassified	21. No. of Pages 53	22. Price n/a

Automated Real-Time Weather Detection Systems Using Artificial Intelligence

Mohamed M. Ahmed, Ph.D., PE

Md Nasim Khan, Ph.D., PE

Anik Das, Ph.D., PE

Department of Civil and Architectural Engineering and Construction Management

University of Wyoming

1000 E. University Ave, Laramie, WY

July 2022

Acknowledgments

The funding for this study was provided by the WYDOT to the Mountain-Plains Consortium (MPC). All statements and opinions presented in this report are the sole responsibility of the authors and may not necessarily reflect those of WYDOT.

Disclaimer

The contents of this report reflect the views of the authors, who are responsible for the facts and the accuracy of the information presented. This document is disseminated under the sponsorship of the Department of Transportation, University Transportation Centers Program, in the interest of information exchange. The U.S. Government assumes no liability for the contents or use thereof.

NDSU does not discriminate in its programs and activities on the basis of age, color, gender expression/identity, genetic information, marital status, national origin, participation in lawful off-campus activity, physical or mental disability, pregnancy, public assistance status, race, religion, sex, sexual orientation, spousal relationship to current employee, or veteran status, as applicable. Direct inquiries to: Vice Provost, Title IX/ADA Coordinator, Old Main 201, 701-231-7708, ndsuoaaa@ndsu.edu.

ABSTRACT

Adverse weather has long been recognized as one of the major causes of motor vehicle crashes due to its negative impact on visibility and road surface. Providing drivers with real-time weather information is therefore extremely important to ensure safe driving in adverse weather. However, identification of road weather and surface conditions is a challenging task because it requires the deployment of expensive weather stations and often needs manual identification and/or verification. Therefore, the primary objective of this research was to develop cost-effective systems capable of providing accurate weather and surface conditions in real time. First, a trajectory-level weather detection system was developed using only a single video camera mounted on the dashboard of the participant vehicles. Two texture-based features, histogram of oriented gradient (HOG) and local binary pattern (LBP), were extracted from images and used as classification parameters to train the weather detection models using several machine learning classifiers, such as gradient boosting (GB), random forest (RF), and support vector machine (SVM). In addition, a unique multilevel model, based on a hierarchical structure, was also proposed to increase detection accuracy. Evaluation results revealed that the multilevel model provided an overall accuracy of 89.2%, which is 3.2%, 7.5%, and 7.9% higher compared with the SVM, RF, and GB model, respectively, using the HOG features. Considering the LBP features, the multilevel model also produced the best performance with an overall accuracy of 91%, which is 1.6%, 8.6%, and 9% higher compared with the SVM, RF, and GB models, respectively. Afterward, the existing roadside webcams in Wyoming were utilized to develop several robust weather and surface detection systems by applying advanced deep learning techniques. Most U.S. departments of transportation (DOTs), including Wyoming DOT (WYDOT), have installed roadside webcams mostly for operational awareness. This study leveraged these easily accessible data sources to develop affordable automatic road weather and surface condition detection systems. The developed detection models are focused on three weather conditions: clear, light snow, and heavy snow; and three surface conditions: dry, snowy, wet/slushy. Several pre-trained convolutional neural network (CNN) models, including AlexNet, GoogLeNet, and ResNet18, were applied with proper modification via transfer learning to achieve the classification tasks. The best performance was achieved using ResNet18 architecture with an unprecedented overall detection accuracy of 97% for weather detection and 99% for surface condition detection. The proposed study has the potential to provide more accurate and consistent weather information in real time that can be made readily available to be used by road users and other transportation agencies. The proposed models could also be used to generate temporal and spatial variations of adverse weather for proper optimization of maintenance vehicle routes and time.

TABLE OF CONTENTS

1. INTRODUCTION.....	1
2. STATE-OF-THE-ART WEATHER DETECTION METHODS	5
2.1 Weather Detection Using Fixed Sources	5
2.2 Weather Detection Using Open-Source Internet Images	5
2.3 Weather Detection Using Moving Sources.....	6
3. DATA COLLECTION, DATA PROCESSING, AND IMAGE ANNOTATION	7
3.1 Trajectory-Level Data	7
3.2 Data from Webcams Along the Interstates in Wyoming	9
4. METHODOLOGY.....	13
4.1 Machine Learning for Trajectory-Level Weather Detection.....	13
4.1.1 Histogram of oriented gradient (HOG) feature.....	13
4.1.2 Local binary pattern (LBP) feature	15
4.1.3 Classification algorithm	15
4.1.4 Development of the multilevel model	16
4.2 Deep Learning for Weather and Surface Detection from Webcams.....	19
4.2.1 Learning algorithm	19
4.2.2 Development of weather and surface condition detection model	21
4.2.3 Performance evaluation of the detection models	21
5. RESULTS AND DISCUSSIONS	23
5.1 Trajectory-Level Weather Detection.....	23
5.1.1 Preliminary investigation of the extracted features	23
5.1.2 Hyper-parameter tuning	25
5.1.3 Performance of the weather detection model using HOG feature	27
5.1.4 Performance of the weather detection model using LBP feature.....	28
5.1.5 Comparison of the computational cost of the weather detection models.....	29
5.1.6 Effect of the number of features on multilevel weather detection	30
5.2 Weather and Surface Condition Detection from Webcams	31
5.2.1 Training and validation	31
5.2.2 Performance evaluation	34
5.2.3 Comparison of the proposed pre-trained CNN with traditional machine learning models	37
6. CONCLUSIONS AND RECOMMENDATIONS	39
7. REFERENCES	41

LIST OF TABLES

Table 1. Criteria for Image Annotation.....	8
Table 2. Criteria For Identifying Weather and Surface Conditions During Image Annotation ...	11
Table 3. Summary Statistics of the Image Datasets form Webcams	12
Table 4. Summary Statistics of the Image Datasets.....	18
Table 5. Performance and Training Time of the Detection Models of the Multilevel Structure..	19
Table 6. Multivariate Analysis of Variance (MANOVA) Test for HOG and LBP Features	25
Table 7. Performance of the Weather Detection Models Using HOG Features	28
Table 8. Performance of the Weather Detection Models Using LBP Features	29
Table 9. Computational Cost of the Weather Detection Models	30
Table 10. Effect of the Number of Features on Multilevel Weather Detection.....	31
Table 11. Model Performance During Validation	34
Table 12. Performance Measures of The Pre-Trained Weather and Surface Condition Detection Models.....	35
Table 13. Detection Summary of the Trained Weather and Surface Condition Detection Model Using ResNet18	36
Table 14. Comparison of the Pre-Trained CNNs with the Traditional Machine Learning Models	38

LIST OF FIGURES

Figure 1. Webcam Locations in Wyoming Road Network	2
Figure 2. Sample Images in Adverse and Clear Weather Collected from Webcams in Wyoming	3
Figure 3. Sample Images of Weather Conditions.....	9
Figure 4. Webcam Locations on Interstate-80, Wyoming	10
Figure 5. Sample Images of Weather Conditions.....	11
Figure 6. Sample Images of Road Surface Conditions	12
Figure 7. Extraction of HOG Feature from Image	14
Figure 8. Extraction of LBP Feature from Image	15
Figure 9. Multilevel Weather Detection based on Categorical Structure.....	17
Figure 10. The Architecture of Convolutional Neural Network (CNN) (29)	20
Figure 11. Comparison of HOG and LBP Features between Different Weather Conditions	24
Figure 12. Parameter Tuning of the Weather Detection Model for Image Dataset-1	26
Figure 13. Training Progress of Weather Detection Models, a) AlexNet, b) GoogLeNet, c) ResNet18.....	32
Figure 14. Training Progress of Surface Condition Detection Models, a) AlexNet, b) GoogLeNet, c) ResNet18	33
Figure 15. Confusion Matrix of the Trained Models Using ResNet18, a) Surface Condition Detection Model, b) Weather Detection Model.....	36

1. INTRODUCTION

Adverse weather events, such as snow, rain, or fog, can directly impact roadway safety by reducing the visibility and roadway surface friction, negatively affecting the vehicle and driver performance, and potentially increasing required stopping sight distance. The Federal Highway Administration (FHWA) revealed that adverse weather is responsible for around 21% of vehicle crashes, 19% of crash injuries, and 16% of crash fatalities each year in the U.S. (1). Previous studies concluded that weather-related factors could increase traffic fatalities and injuries by 25% and 45%, respectively (2, 3). Several studies have concluded that adverse weather can increase the severity of crashes and involve multiple vehicles (4). Snowy weather conditions can result in a sudden reduction in visibility on the roadways. Moreover, road surface friction could be also reduced significantly during snow and heavy rain. The low surface friction of snow-covered roadways coupled with reduced visibility, fog, and heavy rain could result in dangerous conditions for drivers, making it one of the major causes of motor vehicle crashes. According to the FHWA, approximately 688 fatal crashes, 41,860 injury crashes, and 186,076 property damage only (PDO) crashes occur every year in the U.S. because of snow (1). In addition, many pile-up crashes have occurred in recent years due to the presence of fog, which caused many fatalities, injuries, and property damage. For instance, a multi-vehicle crash due to dense fog occurred on I-94 in Michigan on January 9, 2015, which caused the death of one, injured approximately 23 people, and led to closure of the interstate for more than a day (5). Moore and Copper's study noted that despite a 20% decrease in traffic in thick fog, there was an increase of 16% in the total number of personal injury accidents (6). Another study revealed that crashes occurring in heavy fog tend to involve multiple vehicles (7). Considering rainy weather, it was found that the risk of injury crashes in rainy weather conditions could be two times greater than in clear weather conditions (8). Several studies concluded that crashes increase due to vision obstruction during rainfall by 100% or more (8), while others found more moderate but still statistically significant increases (9). However, in the state of Wyoming, the number of snow-related crashes are particularly significant. In the winter of 2018 alone there were 1,438 snow-related crashes, which resulted in fatalities, extended closures, and significant economic loss (10). This is mainly due to Wyoming's adverse winter weather events (such as low visibility and icy road surfaces from blizzard conditions) and the state's roadway and traffic flow conditions (i.e., a large number of low-traffic-volume rural two-lane highways, and mountainous freeways that have a high percentage of heavy inter- and intra-state freight traffic). In practice, the negative impact of snowy weather on roadway safety can be effectively mitigated through the implementation of various safety countermeasures, such as dynamic message signs (DMS) and variable speed limits (VSL) (11, 12). Nevertheless, these countermeasures require accurate and real-time road surface and weather information to operate effectively and reliably. Therefore, the detection of real-time weather conditions and providing drivers with appropriate warnings are crucial for safe driving during adverse weather conditions, including snow, in Wyoming. This is considered by the Wyoming Department of Transportation (WYDOT) Travel Information Service as a primary task (13). The state of practice of collecting and broadcasting road weather information to travelers has been primarily based on roadside weather stations and road weather information systems (RWIS). Although RWIS can provide various weather data, such as temperature, humidity, wind speed, visibility, and precipitation, these systems are expensive. According to the U.S. Department of Transportation, the average total cost of implementing an RWIS is about \$52,000 per unit (14). Therefore, their widespread implementation might not be feasible. In addition, sensors on the weather stations are usually not mounted at the road surface level. Many weather conditions, such as blowing snow, may reduce visibility only at the road surface level due to the accumulation of snow on the side of the road, especially in mountainous regions. In such cases, the visibility at a higher elevation from weather stations might not represent the actual visibility and road surface conditions. Moreover, these weather stations are location-specific and cannot provide real-time trajectory-level weather data. In comparison, the use of webcams, as well as in-vehicle cameras, tends to be a more cost-effective and reliable alternative, and could be installed where power and communication are available. Also, they can provide road weather conditions, including the surface, unlike RWIS.

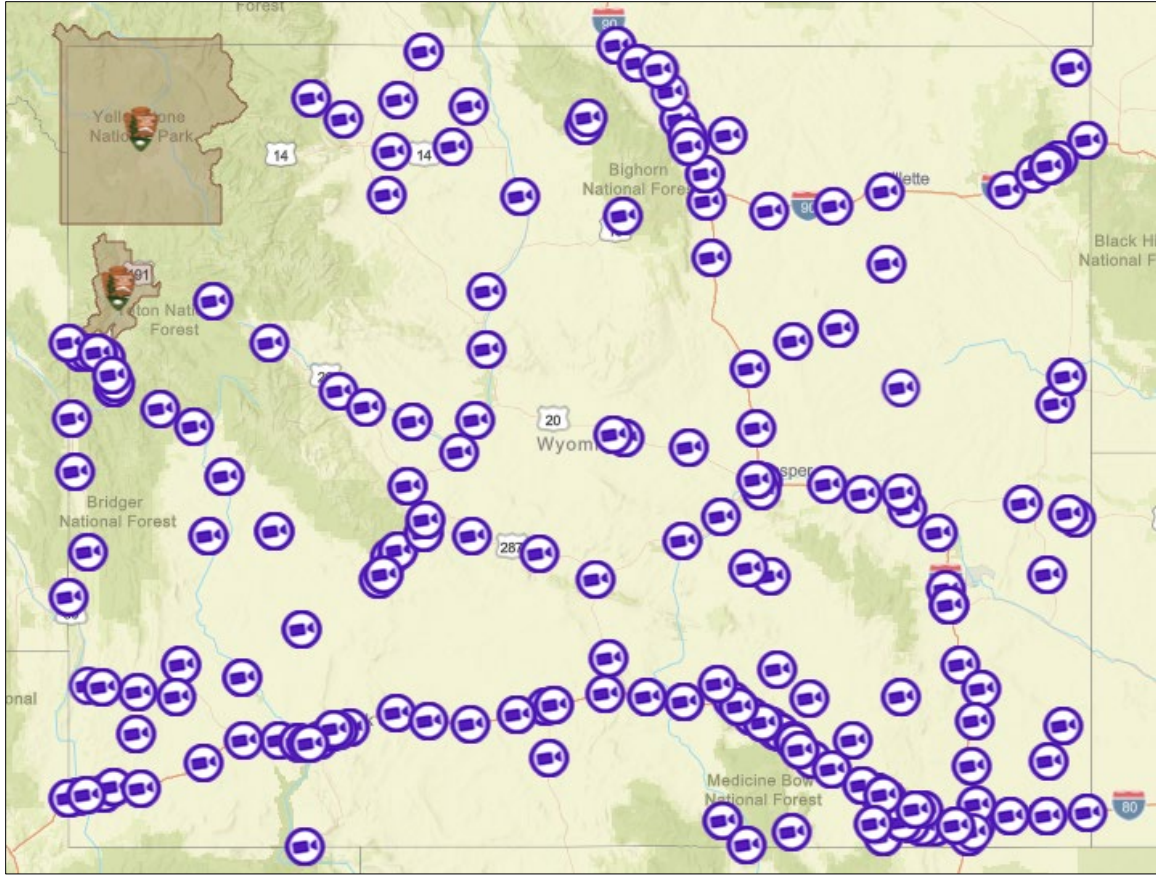
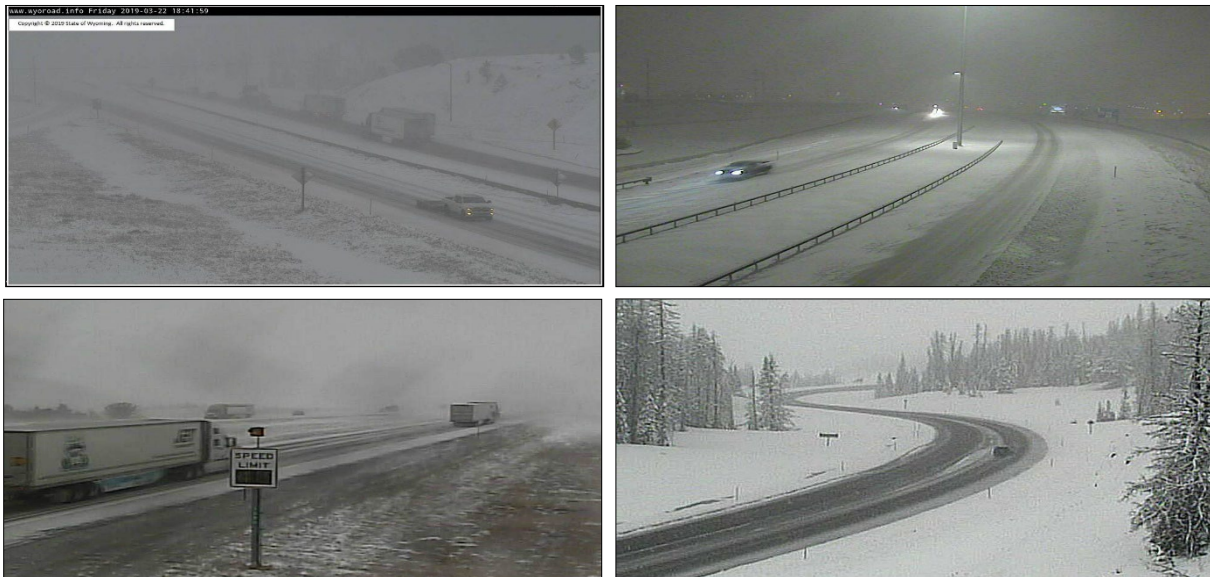


Figure 1. Webcam Locations in Wyoming Road Network

Figure 1 shows the existing webcam locations in the Wyoming Road Network. On the 402-mile Interstate Freeway 80 (I-80) in Wyoming, there are currently 56 webcams with each location having three views of the roadway, including westbound, eastbound, and road surface. The real-time road surface conditions can also be collected, unlike RWIS, since the webcams are also capable of saving the images of the road surface, as shown in Figure 2.



a) Clear Weather



b) Adverse Weather

Figure 2. Sample Images in Adverse and Clear Weather Collected from Webcams in Wyoming

Nevertheless, there are a couple of limitations with using existing webcams or in-vehicle video cameras for real-time weather detection. One of the major limitations is the lack of an accurate and effective video-based automatic weather detection system to process the video images captured by webcams. In addition, under extremely adverse weather conditions, particularly when snowstorms or blizzard events occur, low visibility might impede recognizing road surface level weather. Although in-vehicle cameras can well address this issue, regular vehicles might have to cancel their trips under such weather events, which makes using a regular vehicle for weather data collection not always applicable. As mentioned by the WYDOT snowplow priority plan, the WYDOT snowplow crews will provide service on interstates, high-volume highways, principal arterial and urban routes up to 24 hours a day with a goal of maintaining clear roadways for driving safely at reasonable speeds (10). In current practice, reporting of real-time road

surface winter weather information is mostly based on snowplow drivers. In Wyoming, WYDOT defined nine codes (code #0 to code #8) to represent various weather conditions. Snowplow truck drivers will manually select a code to describe the prevailing surface weather condition of a road segment based on his/her experience and report the code to the WYDOT Traffic Management Center (TMC). However, due to differences in how various weather conditions might be perceived by individual drivers, there might be inconsistencies between reported road weather conditions to TMC. In addition, existing weather codes cannot differentiate the intensity of adverse weather conditions, indicating that the code reported to TMC might not accurately capture the actual adverse weather condition.

The rapid evolution of information technologies (IT) presents opportunities of using machine vision and artificial intelligence (AI) to provide imaging-based automatic detection and analysis of real-time road weather conditions. Machine vision is an integration of a series of technologies, software, and hardware products; it is the science of getting computers to automatically detect patterns in data, and then use the uncovered patterns to predict future data, or to perform other kinds of decision making under uncertainty. The overall machine vision process includes planning the project's details and requirements and then creating a solution. During run-time, the process starts with imaging, followed by automated analysis of the image and extraction of the required information. Given the advantages of machine vision technology, such as real-time processing of road surface level weather conditions, weather detection accuracy, and cost-effectiveness, it has been extensively used in various fields of engineering for image classification, pattern recognition, and text categorization (15).

Considering the limitations of existing WYDOT weather detection systems, and in view of opportunities for the emerging automatic video-image processing technologies, this research aims to develop affordable weather detection systems. Primarily as a proof of concept, the research team used trajectory-level video data to develop weather detection systems capable of detecting seven levels of weather conditions. Afterward, the research team concentrated on the transferability of the research findings in the state of Wyoming. By leveraging webcam images from roadside webcams along Wyoming interstates, several robust models were developed that can provide an accurate estimation of weather conditions in real time. To fulfill the research tasks, a variety of advanced image processing, AI-based techniques, machine and deep learning models, and data mining applications were extensively utilized.

2. STATE-OF-THE-ART WEATHER DETECTION METHODS

Different approaches to image-based weather detection can be found in the literature. Previous studies have developed weather detection models based on different data sources and relying on various image processing techniques, computer vision algorithms, and advanced modeling techniques, including machine learning and deep learning. The related works have been described in the following sections under three broad categories based on the source of data.

2.1 Weather Detection Using Fixed Sources

Many studies have used fixed data sources, such as road weather information systems (RWIS) and closed-circuit television (CCTV), to detect road weather and surface conditions. A study by Jonsson proposed a weather detection system based on the sensor data from RWIS combined with camera images. This study used principal component analysis (PCA) to separate six road conditions, including dry, wet, snow, icy, and snowy with wheel tracks (16). Another study by the same author using the same data sources developed a weather detection algorithm capable of identifying dry, wet, snowy, and icy road conditions with an impressive detection accuracy ranging from 91% to 100%. A study by Carrillo et al. also used similar data sources to develop a surface condition detection system utilizing several pre-trained deep learning models to detect bare, partial snow-covered, and full snow-covered pavement (17). Based on RWIS camera images, the study by Pan et al. leveraged several pre-trained convolutional neural networks (CNN) to detect four road conditions: bare, partially snow-covered, fully snow-covered, and not recognizable. This study achieved an accuracy of more than 97% using the ResNet50 architecture (18). Another study proposed a framework based on CCTV images to detect different situations, such as raining and non-raining scenes, daylight, and nighttime scenes, crowded and non-crowded traffic, and wet and dry roads. This study applied pre-trained neural network models via transfer learning and found an exact match ratio of 0.84. Lee et al. analyzed the colors and edge patterns of the CCTV video to detect sunny, rainy, and cloudy conditions and achieved an overall accuracy of about 86% (19). Another study by the same research group developed an algorithm to estimate the amount of rainfall based on clustering techniques and found an accuracy of 80% (20). A study by Babari et al. proposed a visibility estimation method based on gradient magnitude using roadside highway CCTV images and estimated the visibility with 30% error (21). Another study utilized CCTV images to develop road surface detection models with three categories, including clear, rain-wet, and snow. This study utilized a pre-trained CNN architecture, named VGG16, and achieved an overall detection accuracy of 77% (22).

2.2 Weather Detection Using Open-Source Internet Images

The use of open-source internet images from various platforms, including Google, Flickr, Pixabay, and Yahoo, to develop weather detection models has also been explored in the literature. Ibrahim et al. proposed a new weather detection model, named WeatherNet, using Google images. The proposed WeatherNet was based on ResNet50 architecture and can detect clear, rain, and snowy weather conditions with an overall accuracy of around 93% (23). Another study prepared a comprehensive image dataset, named Img2Weather, consisting of more than 180,000 images in an attempt to develop a weather detection system capable of classifying five weather types: sunny, cloudy, snowy, rainy, and foggy. This study achieved an accuracy of 70% using the random forest model (24). Guerra et al. also created a dataset by extracting images from various platforms, including Creative Commons, Flickr, Pixabay, and Wikimedia Commons. This study proposed a novel algorithm based on CNN architecture and concluded that the proposed model can detect rain, fog, and snow with an accuracy of 80% (25).

2.3 Weather Detection Using Moving Sources

One of the major limitations of weather detection models based on fixed cameras and/or open-source images from the internet is that they cannot provide trajectory-level weather information at the road surface level. Therefore, many studies have used in-vehicle vision systems to detect weather conditions at road surface level. For instance, Pomerleau developed a weather detection system by estimating the reduction of contrast between consistent road features, such as lane markings, shoulder boundaries, and marks left by leading vehicles. The effectiveness of the weather detection system was tested using simulated fog images, as well as real-time images from in-vehicle cameras, which concluded that the system could identify reduced visibility caused by adverse weather conditions (26). Another study developed a weather detection system based on an in-vehicle vision system and AdaBoost classifier and found that the proposed system can classify sunny, cloudy, and rainy weather conditions with an accuracy of 96%, 89%, and 90%, respectively (27). Khan et al. extracted local binary pattern (LBP) based features from snowy images and used three different classification algorithms to detect snow from an in-vehicle video camera (28). Another study by the same authors utilized the SHRP2 Naturalistic Driving Study (NDS) video data to develop a fog detection model based on various neural network architectures and found an overall detection accuracy of 97% in detecting two levels of fog (29). Qian et al. proposed a weather detection system based on dashcams and found 80% accuracy for clear and snow/ice-covered images and 68% for clear/dry, wet, and snow/ice-covered images (30). Another study used an inexpensive car-mounted video camera to capture images of the road surface at nighttime, and which was subsequently used to develop a detection model. They achieved an accuracy of 96%, 89%, and 96% in recognizing dry, wet, and snowy road conditions, respectively (31). Bronte et al. proposed a real-time fog detection system using an onboard low-cost black and white camera. Their system is based on two clues: estimation of the visibility distance, which is calculated from the camera projection equations, and blurring due to the fog (32). Most of the studies based on in-vehicle cameras or sensors require the presence of a consistent object in front of the vehicle. For instance, a weather detection method described in (26) requires road markings, shoulder boundaries, and tracks left by other vehicles. The fog detection system proposed in (33) requires a distinct object in the image. Some studies also used the horizon (34) and the road edge lines (32) to develop a weather detection system

3. DATA COLLECTION, DATA PROCESSING, AND IMAGE ANNOTATION

3.1 Trajectory-Level Data

The video data used in this study were collected from the SHRP2 NDS. This naturalistic driving study was conducted in six U.S. states from 2010 to 2013. This unique dataset has been used in many studies to improve the roadway safety (35–38). All the vehicles that participated in the study were equipped with a data acquisition system (DAS). Along with other data, the ADAS collected videos of the roadways in the moving direction using a color camera under various roadway and weather conditions (39, 40). NDS trips that occurred in adverse weather were collected from the Virginia Tech Transportation Institute (VTTI) using two unique methods, which were developed by the research team exclusively for effective and accurate data collection in various weather (41, 42). In the first method, complementary weather data from the National Centers for Environmental Information (NCEI) were used. The weather stations were considered as points of interest with an influence radius of five nautical miles to find the locations of the trips of interest that occurred in inclement weather. Similarly, the second method used accident locations as points of interest to identify adverse weather trips.

Once the locations and times of adverse weather were identified, videos of the trips that occurred in these locations were collected from VTTI. Subsequently, all the videos were observed manually to verify specific weather conditions and their intensity, i.e., heavy fog, distant fog, etc. Finally, a total of 217 trips in clear conditions, 204 trips in rain, 172 trips in snow, and 168 trips in foggy weather were considered for further analysis. Subsequently, images were extracted from the videos of the selected trips at a 12-frame-per-minute sampling rate, creating an image database consisting of more than 30,000 images. Subsequently, several pre-processing steps, including resizing and cropping, was done on all the images to maintain consistency. The final size of the images was 250 *200 pixels.

After the extraction of images from the videos, all the images were annotated to one of the seven weather levels: clear, light rain, heavy rain, light snow, heavy snow, distant fog, and near fog. Note that the manual annotation of images was challenging and time-consuming. Therefore, to achieve accurate annotation, several criteria were fixed based on quantitative measures to define the weather categories (43). In addition, the research team was provided with comprehensive training to eliminate any potential subjectivity in the annotation process. The criteria used during the manual annotation process are listed in Table 1, and sample images of different categories are illustrated in Figure 3.

Table 1. Criteria for Image Annotation

Weather	Criteria
Clear	Clear visibility Road signs, markings, and surroundings are visible
Light Rain	Raindrops are visible Dry/slightly wet road surface Clear/moderate visibility Road markings and information on road signs and vehicles ahead could be recognized
Heavy Rain	Raindrops are visible Wet road surface Affected visibility Road markings and information on road signs and vehicles ahead could not be recognized
Light Snow	Snowflakes are visible Little/no snow on the road surface Clear/moderate visibility Road markings and information on road signs and vehicles ahead could be recognized
Heavy Snow	Snowflakes are visible Surface covered with snow Affected visibility Road markings and information on road signs and vehicles ahead could not be recognized
Distant Fog	Road markings and information on road signs could be easily recognized Roadside surroundings and traffic ahead are visible The horizon is undefinable
Near Fog	Only a few road markings in front of the NDS vehicle could be observed Information on the road signs could not be read Roadside surroundings and traffic ahead could not be properly recognized The horizon is undefinable



Figure 3. Sample Images of Weather Conditions

3.2 Data from Webcams Along the Interstates in Wyoming

Image data collection, processing, and proper annotation of the extracted images were challenging as well as time-consuming tasks. In an attempt to create an annotated image dataset in adverse weather conditions, first, five months of video data from January to May of 2019 from 56 webcam locations along Interstate-80 were acquired from WDOT. The webcams in each location captured images of at least three views of the roadways, including westbound, eastbound, and road surface, at five minutes intervals. Figure 4 shows the webcam locations along Interstate-80.

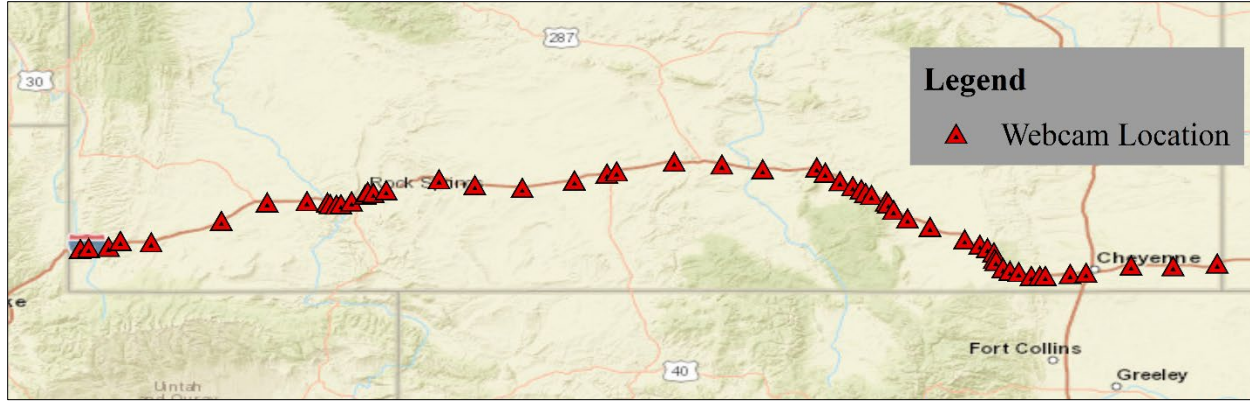


Figure 4. Webcam Locations on Interstate-80, Wyoming

The next challenge was to identify the days with snowy weather conditions from each of these locations. In order to overcome this challenge, historical weather data from the National Climatic Data Center (NCDC) have been utilized and a total of 21 days with minimum snowfall of 0.1 inches have been identified and subsequently used to prepare a database consisting of images under different weather and road surface conditions. Once images were extracted, all the images were normalized, resized, and cropped to maintain consistency, which resulted in an image size of 700 *400 pixels. Subsequently, images with poor visibility were eliminated from the database. During the annotation process, the research team noticed a few easily fixable problems with the extracted images. First, in some locations, the lighting conditions were not adequate to capture good quality images that can be appropriately used for training machine learning models. The use of night vision infrared radiation (IR) cameras or the placement of streetlights next to the webcam locations can potentially solve this issue. The next problem was the accumulation of snow on some of the camera lenses, especially after heavy snowfall. The placement and type of cameras are important in this regard. If wide-angle cameras pointing at the roadways are placed at an angle less than 45 degrees, then it is likely that the snow will slide from the lens and the cameras will be able to capture good quality images. Another potential solution could be the use of a small solar-powered heating system to melt the snow accumulated on the camera lens.

After normalization, the eastbound and westbound images were filtered and annotated manually into three categories: clear, light snow, and heavy snow. The annotation was done based on several qualitative criteria, including the amount of snowfall, visibility of the horizon, roadside surroundings, and road signs. The criteria used for annotating weather conditions are provided in Table 2. The selection of two levels of snow was adopted from a range of literature and in agreement with Highway Capacity Manual (HCM) (28, 37, 43). Similarly, the surface images were separated and grouped into three categories, dry, snowy, and wet/slushy, based on several criteria, as listed in Table 2.

To develop a reliable detection model based on machine vision, proper and accurate annotation is extremely important. Therefore, to ensure accurate annotation free from subjectivity and bias, each image was annotated by more than one observer; subsequently, using the weighted value, the final label was assigned to each image. During the annotation process, all image observers were provided with comprehensive training and sample images in each weather and surface category. Figure 5 and Figure 6 show sample images of each weather and surface category captured at the same locations.

Table 2. Criteria For Identifying Weather and Surface Conditions During Image Annotation

	Category	Criteria
Weather Condition	Clear	• Clear visibility • Road surroundings (e.g., guardrails, delineators, signs, markings, etc.) and traffic are visible • Horizon is visible
	Light Snow	• Snowflakes are visible • Little/no snow on the road surface • Clear/moderate visibility • Road surroundings and traffic are visible to some extent • Horizon undefinable
	Heavy Snow	• Snowflakes are visible • Surface covered with snow • Affected visibility • Road surroundings and traffic are visible to some extent • Horizon undefinable
Surface Condition	Dry	• Surface free of precipitation • No/little snowfall
	Snowy	• Surface partially or fully covered with snow • Moderate/heavy snowfall • Temperature below the freezing point
	Wet/Slushy	• Wet surface or covered with slush • Little/moderate snowfall • Temperature above the freezing point



Figure 5. Sample Images of Weather Conditions

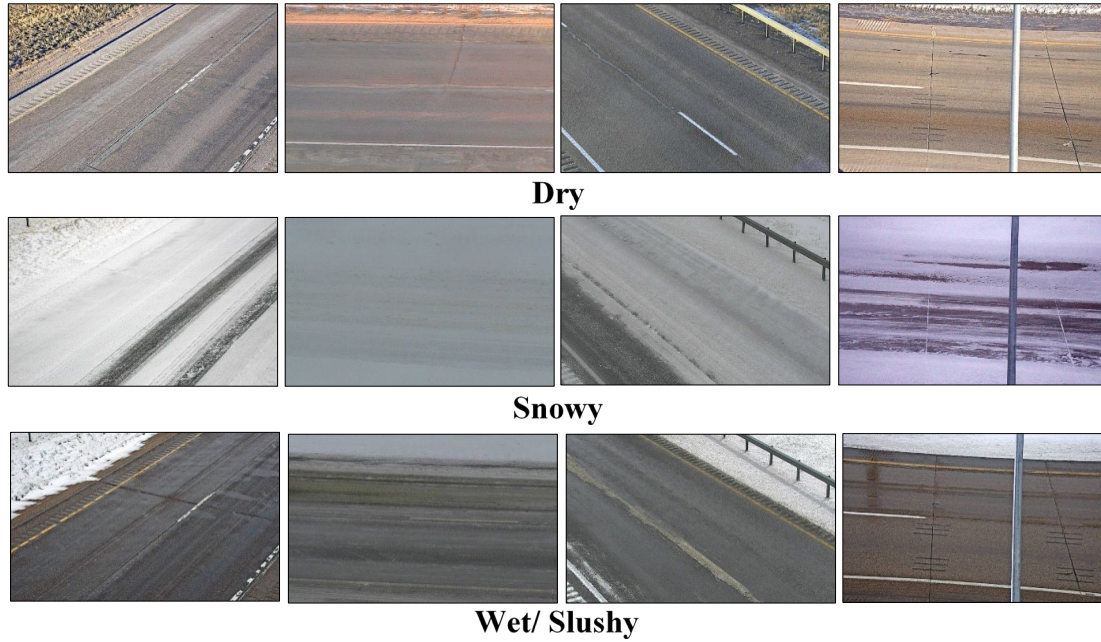


Figure 6. Sample Images of Road Surface Conditions

After the data reduction and processing, the final image datasets for weather detection consisted of 15,000 images, with 5,000 images in each weather category. For training and validation, 80% of the total images (12,000 images with 4,000 images in each weather category) were utilized and the remaining 20% (3,000 images with 1,000 images in each weather category) were used to test the performance of the weather detection models. Note that the test images were never used during training and validation. The surface condition detection dataset also consisted of 15,000 images with 5,000 images in each surface condition category with similar kinds of splits. A summary of the image data used in this study is listed in Table 3.

Table 3. Summary Statistics of the Image Datasets from Webcams

	Total	Categories	Training and Validation	Testing
Weather Detection	15,000	Clear	4000	1,000
		Light Snow	4000	1,000
		Heavy Snow	4000	1,000
		<i>Total</i>	<i>12,000</i>	<i>3,000</i>
Surface Condition Detection	15,000	Dry	4000	1,000
		Snowy	4000	1,000
		Wet/ Slushy	4000	1,000
		<i>Total</i>	<i>12,000</i>	<i>3,000</i>

4. METHODOLOGY

4.1 Machine Learning for Trajectory-Level Weather Detection

Machine learning techniques are currently being used widely in various fields, including transportation engineering for detection, pattern recognition, and safety purposes (44–46). Any traditional machine learning approach usually requires two major steps: feature extraction and classification algorithm. The feature extraction task is usually achieved by selecting interesting points or objects in the images. However, for weather detection purposes, this approach is impractical because under different weather conditions, the same points or objects can be present. Therefore, it is important to select proper image features capable of providing appropriate information regarding weather conditions that can later be used to train detection models using machine learning algorithms. In this study, two image features, including HOG and LBP, have been used to develop weather detection models.

4.1.1 Histogram of oriented gradient (HOG) feature

The sharpness of an image can be highly affected by weather conditions because different weather can cause a different degree of blur. In clear weather, the images are usually sharper, and it is more likely to have a sudden change in pixel intensity in an image. Conversely, in adverse weather, images are always blurred, and more uniform pixel intensity is expected. This change in sharpness due to adverse weather can effectively be measured using HOG (47). The HOG was first introduced by Dalal et al. (48) and can be used for pattern recognition, object detection, and image classification (49, 50). The goal of this method is to describe an image by a set of local histograms. These histograms represent the occurrences of gradient orientation in localized portions of an image.

The first step of computing the HOG feature is to calculate the gradients of the image. The image gradient is defined as a directional change in pixel intensity on both the x-axis and y-axis. The gradient vector of a pixel at location (x, y) can be described using Equation 1.

$$\nabla f(x, y) = \begin{bmatrix} g_x \\ g_y \end{bmatrix} = \begin{bmatrix} \frac{\partial f}{\partial x} \\ \frac{\partial f}{\partial y} \end{bmatrix} = \begin{bmatrix} f(x+1, y) - f(x-1, y) \\ f(x, y+1) - f(x, y-1) \end{bmatrix} \quad (1)$$

Where $f(x, y)$ is the pixel intensity at coordinates x and y ; g_x and g_y are the gradient in x and y direction, respectively. The magnitude, $M(x, y)$ and phase, $\theta(x, y)$ of the gradient, then can be calculated using the following equations.

$$M(x, y) = \sqrt{g_x^2 + g_y^2} \quad (2)$$

$$\theta(x, y) = \arctan \frac{g_y}{g_x} \quad (3)$$

Where g_x and g_y are the gradient in x and y direction, respectively.

The next step is to divide each image into multiple cells, as illustrated in Figure 7a. For this study, a cell size of 50×50 pixels has been used. Since the size of the image is 250×200 pixels, using a 50×50 pixels cell produced five cells in the horizontal direction and four cells in the vertical direction. Note that each cell has a total of $50 \times 50 = 2,500$ -pixel values and the gradient

of each pixel contains two values, representing magnitude and direction, as calculated previously using Equation 2 and Equation 3. In each cell, the magnitude of these 2,500 pixels was then cumulatively added into five bins corresponding to angles 0, 36, 72, 108, and 144 to form the histogram of gradient as depicted in Figure 7b.

Once the histogram of the gradient was obtained, a block normalization technique with a block size of 2×2 cells was applied to each image, as illustrated in Figure 7c. The normalization was necessary for eliminating any potential effect of lighting variations (48). Since a 2×2 block contained four histograms with five bins in each histogram, concentrating the block produced a vector of size 20×1 . A normalization factor was then calculated for the vector containing all histograms of a given block using “L2 norm.” The “L2 norm” is the square root of the sum of the squared vector values and can be described using Equation 4 (51, 52).

$$f = \frac{v}{\sqrt{\|v\|^2 + e^2}} \quad (4)$$

Where v is the vector containing all histograms in a block, f is the normalization factor, e is a small regularization constant, and $\|v\|$ is the vector norm.

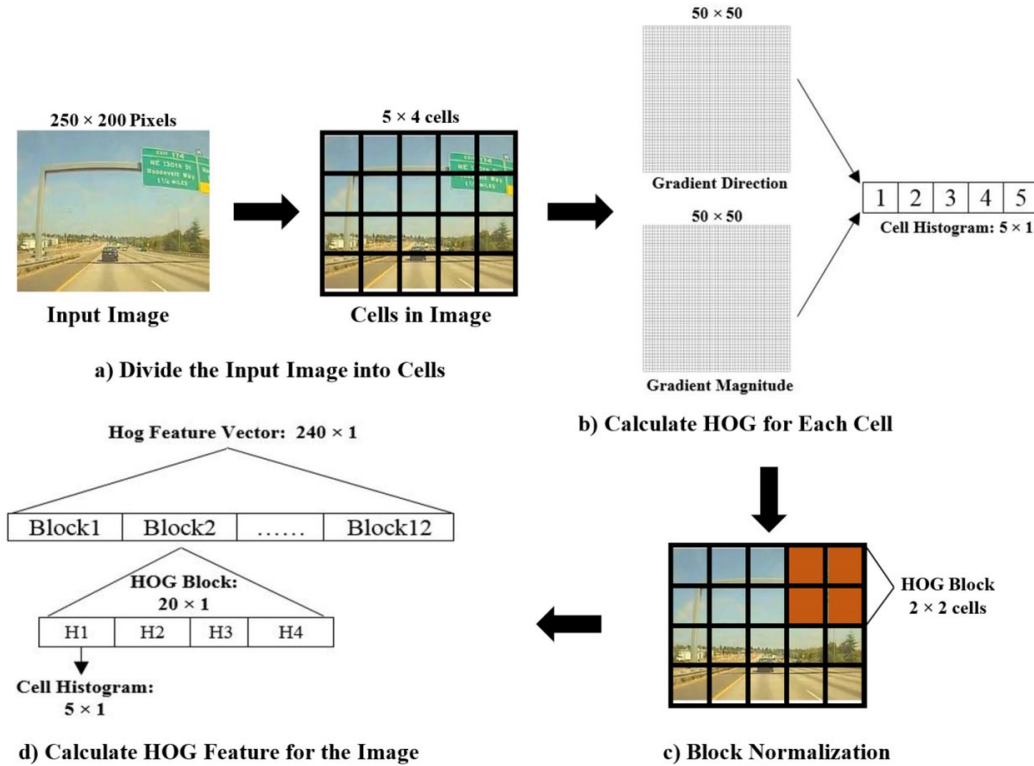


Figure 7. Extraction of HOG Feature from Image

After the normalization, all the vectors from each block of the images were concentrated into one giant vector, as seen in Figure 7d. Since a block of size 2×2 cells can have 4 horizontal and 3 vertical positions in a 5×4 cells image (250×200 pixels), a total of $4 \times 3 = 12$ positions of the block were possible. As mentioned earlier each block was represented by a 20×1 vector. Therefore, once all the blocks were concentrated, the HOG feature vector was found to have $20 \times 12 = 240$ entries.

4.1.2 Local binary pattern (LBP) feature

LBP can effectively be used to extract texture-based image features for classification and pattern recognition purposes and was proposed in (53). LBP compares each pixel of an image to its neighborhood pixels to express the local variation of image texture. The LBP was initially developed considering a fixed 3×3 window. However, a modified version of LBP with two parameters was used in this study to capture the texture variation in different scales. The first parameter represents the number of pixels around the center pixel and was denoted by P , and the second parameter represents the radius of the circle and was denoted by R . The LBP features can be defined using Equation 5.

$$LBP_{P,R} = \sum_{p=0}^{P-1} s(g_p - g_c) 2^p \quad (5)$$

$$s(x) = \begin{cases} 1, & x \geq 0 \\ 0, & x < 0 \end{cases}$$

Here, P is the number of pixels, R is the radius, and g_p and g_c is the gray-level intensity of the surrounding and center pixel, respectively.

In this study, 8 pixels ($P = 8$) around the center pixel and a radius of 1 ($R = 1$) were used, as seen in Figure 8. First, the differences in the grey-level intensity between the center pixel and neighboring pixels were calculated; then based on the values, each pixel was assigned with either 0 for negative value or 1 for positive value, which ultimately produced a binary pattern of the center pixel. Similarly, binary patterns of all the pixels of an image were calculated, and then the patterns were categorized into two levels: uniform and non-uniform. A pattern was defined as uniform if it has a maximum of two 0-1 or 1-0 transitions. For instance, 10000000 can be considered a uniform pattern since it has only one 1-0 transition. Similarly, 11001001 can be considered a non-uniform pattern since it has four 0-1 or 1-0 transitions. Since $LBP_{8,1}$ was used, 256 patterns were extracted with 58 uniform and 198 non-uniform patterns. Separate labels were assigned to each uniform pattern, and a single level was assigned to all the non-uniform patterns; therefore, a total of 59 different labels were created (54, 55).

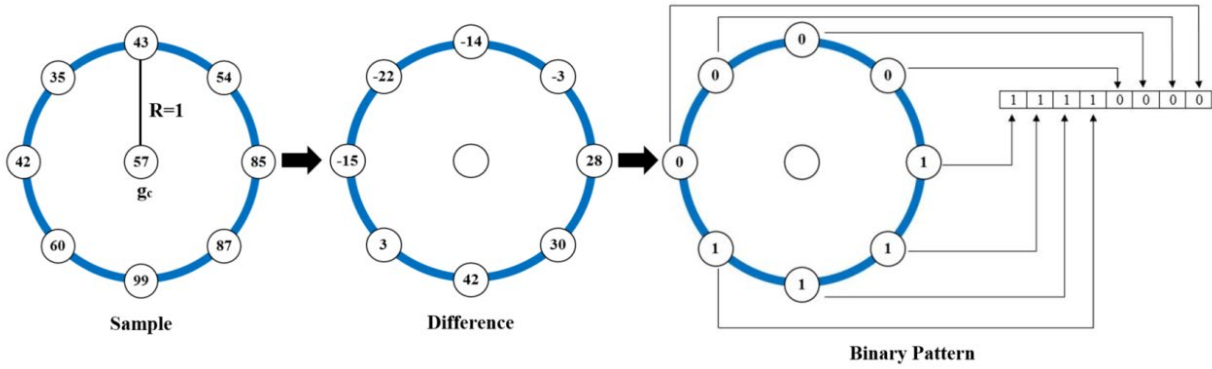


Figure 8. Extraction of LBP Feature from Image

4.1.3 Classification algorithm

To train the weather detection model based on the HOG and LBP features, this study first considered three widely used machine learning classifiers: support vector machine (SVM), random forest (RF), and gradient boosting (GB). This study then also developed a multilevel detection model, where instead of using a single model for all weather conditions combined, four different machine learning classifiers were merged in a hierarchical structure to get better detection accuracy.

SVM is a supervised machine learning method based on the concept of optimal hyperplane and can be used for classification. It utilizes the kernel function to transform data, which then can be used to construct an optimal boundary, also known as a hyperplane, between the possible outputs. The points closer to the hyperplane are defined as support vectors. They are the points that are most difficult to classify and used to maximize the margin of the classifier. In order to get better classification results, the margin between the data points and the hyperplane needs to be maximized using a cost function. Two main parameters can be tuned during the training process: the regularization parameter, also known as C parameter, and gamma parameter. The C parameter controls the amount of allowable misclassification during optimization. For a large value of C , the optimization will look for a hyperplane with a smaller margin capable of separating all the training points correctly; whereas, for a small value of C the optimizer will look for a hyperplane with a larger margin, even if the hyperplane produces some misclassification. The gamma parameter controls the influence area during a single training. For high gamma, only the nearby points from the plausible line are considered in the calculation for creating the hyperplane. On the other hand, with low gamma, points far from the plausible line are also considered (56, 57).

RF is an ensemble machine learning method that can be used to train models for both classification and regression. RF algorithm consists of numerous randomly selected decision trees. Two types of randomness are generally introduced into the model: each tree is created from a randomly selected sample of the training data, and at each node the variables are randomly selected to produce the best split. The randomness increases diversity in the model, leading to a more accurate overall prediction. To classify a new point, all the decision trees within the forest produce a prediction, and the RF algorithm selects the prediction with the largest number of votes. One of the advantages of using RF is its ability to reduce the risk of overfitting (58).

GB is an ensemble learning technique, which produces a prediction model by converting weak learners into strong learners through the sequential adjustment of the weights based on errors from the previous iteration. In GB, the individual models are not based on a random selection of data; rather they are built sequentially where, in each successive iteration, the weights are adjusted by putting more weights on observations that are hard to predict and reducing the weights of those that are easy to predict. In GB, the gradient descent is used to minimize the loss function at every iteration. Note that the loss function is a measure of the prediction ability of a model. The hyperparameters of the GB are the learning rate, reduction of the learning rate (also known as shrinkage), boosting iteration, and decision-tree-related parameters (59).

4.1.4 Development of the multilevel model

In an attempt to improve the detection performance, the research team devised an algorithm, where instead of using a single model for all the weathers, four different machine learning classifiers were merged into a hierarchical structure. The steps of the proposed multilevel weather detection model are described as follows and illustrated in Figure 9.

- 1) Train 4 separate detection models for each level: L1-based on image database-2 with four categories (clear, rain, snow, and fog); L2R-based on image dataset-3 with two categories (light rain and heavy rain); L2S-based on image dataset-4 with two categories (light snow and heavy snow); L2F-based on image dataset-5 with two categories (distant fog and near fog). For each of these levels, SVM, RF, and GB were trained separately, and the best-performing model was selected.
- 2) Input the test image into the L1 model, and get the temporary weather category, C_{temp} .

- 3) If C_{temp} is clear, the final category is clear; if C_{temp} is rainy, pass the input image to the L2R model to get the final rain category; if C_{temp} is snowy, pass the input image to the L2S model to get the final snow category; and if C_{temp} is foggy pass the input image to the L2F model to get the final fog category.

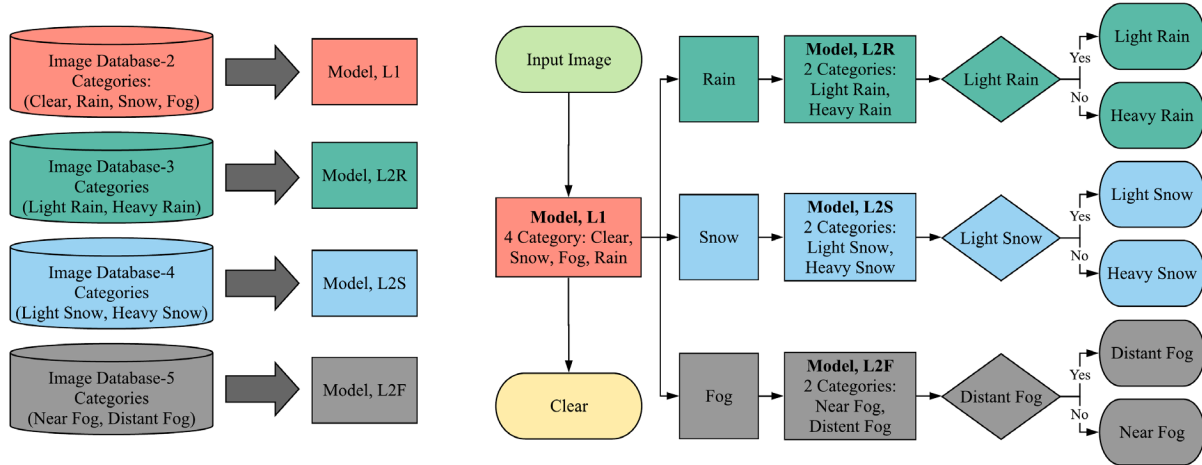


Figure 9. Multilevel Weather Detection based on Categorical Structure

In this study, five separate datasets were used to train, validate, and test different weather detection models. For each image dataset, 80% of the images were used to train and validate the detection models, and the rest of the data were used to test the performance of the calibrated models. Note that the test data were not used during training and validation. In addition, 5-fold cross-validation was applied to validate and tune the hyperparameters of the models. The image dataset-1, which consisted of seven categories with 2,500 images in each category, was used to train and validate the SVM, RF, and GB models, as well as to test the accuracy of the SVM, RF, GB, and multilevel models. The rest of the datasets were used to train, validate, and test the detection models inside the multilevel weather detection structure. Image dataset-2, image dataset-3, image dataset-4, and image dataset-5 were used to train, validate, and test the L1, L2R, L2S, and L2F detection models, respectively. The summary statistics of the image data used in this study are shown in Table 4.

Table 4. Summary Statistics of the Image Datasets

Datasets	Purpose	Weather Categories	For Training and Validation		For Testing	
			Number of Images	Equivalent video Duration (min)	Number of Images	Equivalent video Duration (min)
Image Dataset-1 (7 categories)	- To train and validate the SVM, RF, and GB model - To test the SVM, RF, GB, and multilevel model	Clear	2,000	166.67	500	41.67
		Light Rain	2,000	166.67	500	41.67
		Heavy Rain	2,000	166.67	500	41.67
		Light Snow	2,000	166.67	500	41.67
		Heavy Snow	2,000	166.67	500	41.67
		Near Fog	2,000	166.67	500	41.67
		Distant Fog	2,000	166.67	500	41.67
		Total	14,000	1166.67	3,500	291.67
Image Dataset-2 (4 Categories)	To train validate, and test the L1 of the multilevel model	Clear	4,000	333.33	1,000	83.33
		Rain	4,000	333.33	1,000	83.33
		Snow	4,000	333.33	1,000	83.33
		Fog	4,000	333.33	1,000	83.33
		Total	16,000	1333.33	4,000	333.33
Image Dataset-3 (2 Categories)	To train validate, and test the L2R of the multilevel model	Light rain	2,000	166.67	500	41.67
		Heavy Rain	2,000	166.67	500	41.67
		Total	4,000	333.33	1,000	83.33
Image Dataset-4 (2 Categories)	To train validate, and test the L2S of the multilevel model	Light Snow	2,000	166.67	500	41.67
		Heavy Snow	2,000	166.67	500	41.67
		Total	4,000	333.33	1,000	83.33
Image Dataset-5 (2 Categories)	To train validate, and test the L2F of the multilevel model	Distant Fog	2,000	166.67	500	41.67
		Near Fog	2,000	166.67	500	41.67
		Total	4,000	333.33	1,000	83.33

As mentioned earlier, the proposed multilevel model consists of four independent detection models (e.g., L1, L2R, L2S, and L2F) in a hierarchical structure. For each of these models, SVM, RF, and GB classifiers were trained separately, and the best-performing models were selected for the final weather detection model. Considering the HOG-based features, the SVM produced the best performance in terms of detection accuracy and computational cost for all four models, as shown in Table 5. For instance, the performance of the SVM classifier based on HOG features for the L1 model was significantly better than the RF and GBT models with the highest detection accuracy of 93.3% and the lowest training time of 122.3 seconds. The LBP-based models also produced similar results, where the SVM classifier yielded the best performance in terms of overall accuracy and computational cost for L1, L2R, and L2S models. However, considering the L2F model, the RF classifier provided the highest detection accuracy of 95.3%, although it required more time to train compared with the other models. Since the multilevel structure combined the L1, L2R, L2S, and L2F models, the final training time of the multilevel models were calculated by adding the sum of the training times of the best-performing models, which were around 138 seconds and 53 seconds for HOG- and LBP-based features, respectively.

Table 5. Performance and Training Time of the Detection Models of the Multilevel Structure

Feature Extraction Method	Model Name	Categories	Classifier	Overall Accuracy (%)	Training Time (s)
HOG (240 Features)	L1	Clear, Rain, Snow, Fog	SVM	93.3	122.3
			RF	88.3	272.8
			GBM	88.7	372.3
	L2R	Light Rain, Heavy Rain	SVM	93.8	5.8
			RF	91.7	37.2
			GBM	91.9	43.2
	L2S	Light Snow, Heavy Snow	SVM	95.1	5.4
			RF	93.0	35.7
			GBM	93.1	41.9
	L2F	Near Fog, Distant Fog	SVM	96.9	5.1
			RF	95.5	34.2
			GBM	95.8	41.9
LBP (59 Features)	L1	Clear, Rain, Snow, Fog	SVM	91.3	39.1
			RF	89.8	67.0
			GBM	90.1	59.9
	L2R	Light Rain, Heavy Rain	SVM	93.3	2.1
			RF	91.3	9.8
			GBM	91.0	8.72
	L2S	Light Snow, Heavy Snow	SVM	93.3	1.9
			RF	92.2	9.9
			GBM	92.0	8.7
	L2F	Near Fog, Distant Fog	SVM	91.4	2.1
			RF	95.3	9.9
			GBM	94.8	8.6

4.2 Deep Learning for Weather and Surface Detection from Webcams

4.2.1 Learning algorithm

To train and validate the weather and surface detection models, this study leveraged a cutting-edge deep learning technique named convolutional neural network (CNN). CNN was developed specifically for solving image classification problems, and previous studies have revealed that it can provide a high degree of detection accuracy and can outperform other traditional machine learning models (60–62). Similar to all other deep learning models, the architecture of CNN can be broadly categorized into three types of layers: an input layer, hidden layers, and an output layer.

The primary purpose of the input layer is to receive the annotated input images and pass them to the subsequent hidden layers. The majority of the computation occurred at the hidden layers, which can be grouped into three types of layers, including convolutional, rectified linear unit (ReLU), and pooling layer. The convolution layer is the main building block of a CNN and consists of several filters with sizes less than the input images. These filters are moved across the input image in such a way that all the pixels are covered at least once and the dot product between the filter and the input is generated at every special position of the image. The resulting outputs from all the filters are then piled along the depth dimension to get the output of the convolution layer. The main purpose of the convolution layer is to extract features from the input image. While the initial convolution layers extract more generic features, as the network gets deep, the subsequent convolution layers extract more refined features, which are more suited to solve

the classification problem (63). The purpose of the ReLU layer is to perform a threshold operation on each element of the inputs to ensure fast and consistent training. The ReLU layer applies a function that returns the positive values directly and converts the negative values to zero (64, 65). Finally, the pooling layer reduces the amount of information generated from the preceding convolution layer to ensure the passing of only the most essential information to the next layers (29). The output layer usually consists of two types of layers: fully connected layer and softmax layer. The fully connected layer produces an output vector with dimensions equal to the number of output classes, which is then passed to the softmax layer to assign decimal probabilities to each of the output classes (29). The overall architecture of CNN is illustrated in Figure 10.

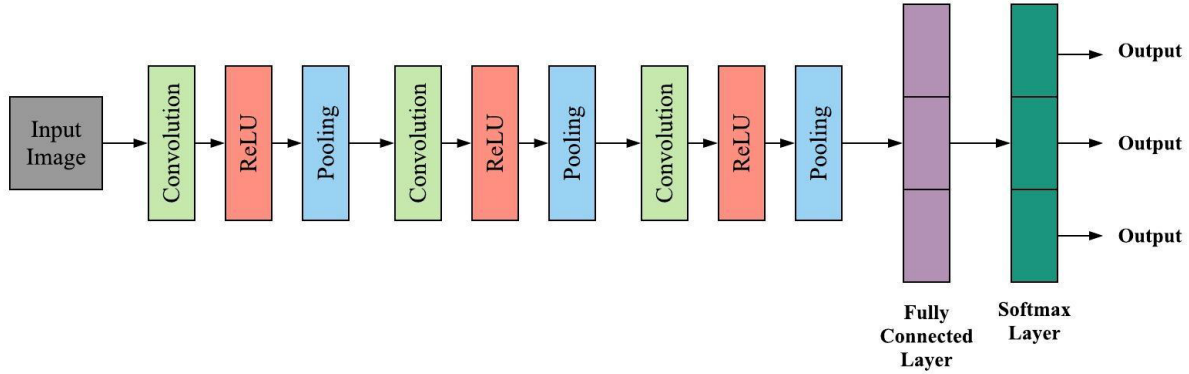


Figure 10. The Architecture of Convolutional Neural Network (CNN) (29)

Training of a CNN model from scratch requires large-scale precisely annotated image data, powerful computational resources, and careful crafting of the neural network architecture. Another approach of training a CNN model is transfer learning, which used the experiences gained from any of the previously trained CNN models. The pre-trained model can provide excellent detection accuracy and can outperform CNN models, which are built on completely new image data with relatively small training samples. Considering the advantages of the transfer learning technique, this study also utilized this method to train the weather and surface condition detection models. Previous studies have experimented with transfer learning using a number of pre-trained CNN models, including AlexNet, VGGNet, GoogLeNet, Inception, xception, SqueezeNet, MobileNet, DenseNet, ResNet18, and ResNet50 (66, 67). For this study, AlexNet, GoogLeNet, and ResNet18 are trained and comparatively evaluated to find the best performing weather and surface condition detection models. These models were selected considering their simple structure, fast training time, the capability of achieving high accuracy, and relatively less requirement for computation power. The following section briefly describes the architecture of the pre-trained model used in this study.

AlexNet, proposed by Krizhevsky et. al. (60), was the first architecture that demonstrated the potential of using CNN in image classification and pattern recognition. AlexNet earned the ImageNet ILSVRC challenge in 2012 by outperforming other models by a significant margin. The architecture of AlexNet consists of five convolutional layers and three fully connected layers. In addition, the architecture also introduced ReLU layers for the first time, which substantially improved the training time of the model. AlexNet was trained using a subset of the ImageNet dataset and can identify 1,000 possible classes. Note that ImageNet is a massive image dataset with around 15 million annotated images and contains more than 20,000 categories (60).

GoogLeNet was developed by a team at Google® with the aim of maximizing the computing resources inside the network (68). The problem of overfitting due to many deep layers was also addressed in this

architecture by introducing filters with multiple sizes capable of operating at the same level. However, its main advantage was the introduction of an “Inception Module” that drastically reduced the number of parameters to only four million compared with 60 million with AlexNet. The GoogLeNet architecture consists of 22 deep layers with nine inception modules arranged linearly. The ends of the inception modules are linked to a global pooling layer to improve the performance in terms of recognition accuracy.

ResNet18 presented an easily optimizable residual learning framework that enables the architecture to train ultra-deep neural networks. Training of deep neural networks is difficult because of the vanishing problem, which causes rapid degradation of accuracy with the increase of more layers in the network. ResNet18 solves this problem by introducing a deep residual learning framework, and therefore can provide a high degree of accuracy compared with other CNN models with similar depth (69). ResNet-18 has a total of 18 layers with serially connected basic blocks in addition to shortcut connections parallel to each basic block (70).

4.2.2 Development of weather and surface condition detection model

As mentioned earlier, to develop the detection models, this study leveraged three pre-trained CNN models, including AlexNet, GoogLeNet, and ResNet18. Different sets of CNN models were developed for weather and surface condition detection. In order to apply the pre-trained models, transfer learning techniques have been utilized, where the last few layers of the pre-trained models have been modified or replaced to achieve the weather and surface condition detection tasks. For AlexNet, the last two layers were removed and replaced with two new layers: a fully connected layer with three classes and an output layer. Similarly, the last three layers of the GoogLeNet were updated with a fully connected layer, a softmax layer, and a classification layer to accommodate the new image categories. Similar strategies were also followed for updating the ResNet18. In addition, the input image size was also modified to meet the requirements of the pre-trained models. Note that AlexNet requires an input image size of 227×227 pixels, whereas both GoogLeNet and ResNet18 require an image size of 224×224 pixels.

The initial hyperparameters of the model were updated by carefully observing the training progress and validation results for different parameters. Note that 80% of the training images were used for training and the remaining 20% were used for validation. The 80%-20% split is widely used in the literature for developing machine learning models and is based on the Pareto Principle (71, 72). In this study, the overall validation accuracy was calculated and reported every 15 iterations.

4.2.3 Performance evaluation of the detection models

The performance of the models was evaluated by applying seven performance indices: sensitivity, false negative rate (FNR), specificity, false positive rate (FPR), accuracy, precision, and F1-score. Previous studies have used these indices to evaluate the quality of the learning algorithm (67, 73, 74). Sensitivity or recall measures the ability of the model to correctly classify an image group. In other words, it represents the degree of accuracy of positive examples and can be defined by Equation 1, where TP and FN refer to true positives and false negatives, respectively, of an image group. TP represents the number of images that are correctly classified, whereas FN is the number of images that are misclassified as other images. In addition, the counterpart of recall is called FNR and can be described using Equation 2.

$$\text{Sensitivity (Recall)} = \frac{TP}{TP+FN} \quad (1)$$

$$FNR = 1 - \text{Recall} \quad (2)$$

Specificity or true negative rate (TNR) corresponds to the proportion of negatives cases that were classified correctly and can be expressed using Equation 3, where TN and FP represent true negative and false positive, respectively, of an image group. TN corresponds to the number of negative cases that are correctly classified, and FP corresponds to the negative cases that are misclassified as positive cases. FPR is the counterpart of specificity and can be defined using Equation 4.

$$Specificity (TNR) = \frac{TN}{TN+FP} \quad (3)$$

$$FPR = 1 - Specificity \quad (4)$$

Accuracy measures the overall ability of detection models to correct classification and is the most widely used matrix to evaluate the performance classification; however, its value could be misleading, especially, for classification problems with imbalanced data, where a particular image group has a significantly higher recall value compared with other groups. The overall accuracy of a model can be defined using the following equation.

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \quad (5)$$

Precision indicates the ability of a model to classify images correctly, with no false prediction, and can be described using Equation 6. In other words, precision represents the predictive power of a model because it represents the degree of correctly identified positive images out of all the predicted positive images (73).

$$Precision = \frac{TP}{TP+FP} \quad (6)$$

In classification problems, a model might have high recall with low precision value and vice versa. The best performance is achieved only when a model has balanced high recall and precision values, which could be measured using an index called F1-score, as described in Equation 7. F1-score represents the harmonic mean of precision and recall, and a high F1-score indicates that the model is balanced with a high recall and precision value.

$$F1 - score = 2 \times \frac{Precision \times Recall}{Precision + Recall} \quad (7)$$

5. RESULTS AND DISCUSSIONS

5.1 Trajectory-Level Weather Detection

The detailed descriptions of the findings from this study have been presented in the following section. First, preliminary investigations of the extracted features have been conducted to examine significant differences among the image groups. Next, the results from the hyper-parameter tuning have been discussed. Afterward, the performance of the trained machine learning models has been described in terms of several performance indices. Subsequently, a comprehensive compression of the computational cost of different models has been provided. Finally, the effect of the number of features on multilevel weather detection has been investigated. The HOG and LBP features were extracted using the Computer Vision Toolbox™ in MATLAB® version 9.8 (R2020a). Once the features were extracted, all analyses were conducted in R® programming language version 3.6.3. R® is an open-source programming software for statistical computations and machine learning modeling. Most of the recent developments in the field of statistics and machine learning are usually available in R® through different packages and can be used on most operating platforms, including Windows, macOS, and Linux. All the machine learning models were trained, validated, and tested using the “caret” package in R®.

5.1.1 Preliminary investigation of the extracted features

As mentioned earlier, 240 HOG features and 59 LBP features were used to develop the weather detection models. Before starting the modeling step, the value of all the features was thoroughly investigated in order to determine if the features can be used as training parameters. According to Figure 11, the average values of both HOG and LBP features are significantly different among the categories. For instance, the HOG features of the clear image group have relatively higher values compared with the other image groups. Similarly, the heavy snow image group has higher values compared with the light snow image group for most of the HOG features. Similar kinds of variations were also found for other image groups

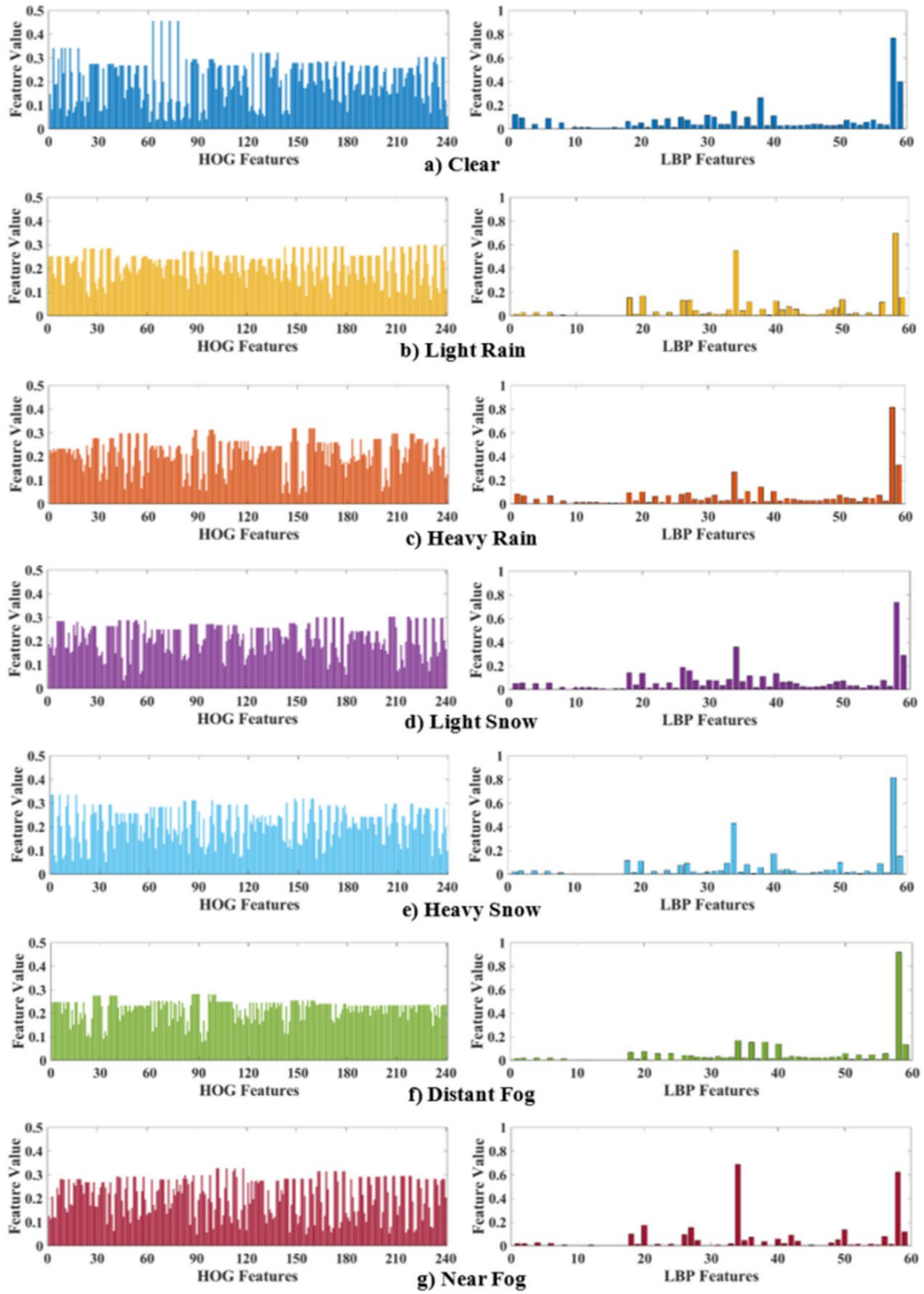


Figure 11. Comparison of HOG and LBP Features between Different Weather Conditions

Multivariate analysis of variance (MANOVA) was used to determine whether there are differences in LBP and HOG features among the image groups. MANOVA is a powerful statistical technique that can simultaneously test the differences in population means of independent variables between multiple dependent variables (Finch, 2005). In this study, weather categories are the dependent variables, whereas 240 HOG features and 59 LBP features are the independent variables. Table 6 provided the MANOVA test results based on four test statistics.

Table 6. Multivariate Analysis of Variance (MANOVA) Test for HOG and LBP Features

		Test statics	F	Hypothesis df	Error df	P-value
HOG Features	Wilks' Lambda	2.9	52.6	1440	82554	< 0.01
	Pillai Trace	0.1	60.3	1440	82511	< 0.01
	Hotelling-Lawley	7.2	69.0	1440	82514	< 0.01
	Roy's Largest Root	2.6	151.62	240	13759	< 0.01
LBP Features	Wilks' Lambda	2.20	294.37	354	179640	< 0.01
	Pillai Trace	0.03	391.35	354	178822	< 0.01
	Hotelling-Lawley	6.47	546.81	354	179600	< 0.01
	Roy's Largest Root	3.96	2007.03	59	29940	< 0.01

Considering the HOG features, the statistical hypothesis can be described as follows:

$$H_0: \mu_1 = \mu_2 = \mu_3 \dots \dots \dots = \mu_{240}$$

$$H_1: \mu_1 \neq \mu_2 \neq \mu_3 \dots \dots \dots \neq \mu_{240}$$

Here, H_0 = null hypothesis, H_1 = alternative hypothesis, μ = population mean.

The Wilks' Lambda test, with $F(1440, 82554) = 52.6$ and $p\text{-value} < 0.01$, indicates that H_0 should be rejected at a 99% confidence level. In other words, the population means of the HOG features are significantly different among the image groups at a 99% confidence level. Other test statistics also rejected the null hypothesis. Similarly, the LBP features were also found to be significantly different among the image groups, as seen from Table 6. Overall, the figures and the statistical tests indicate that the HOG and LBP features are significantly different among the image groups and therefore can be effectively utilized to train the weather detection models.

5.1.2 Hyper-parameter tuning

To identify the best SVM, RF, and GB models using the image features, instead of using the default values, the hyperparameters of these models were tuned. Hyperparameters are specific properties that are used to control various aspects of the machine learning algorithms and can have substantial effects on their complexity and performance (Chicco, 2017). However, there is no straightforward approach to tune the hyperparameters, and tuning mostly relies on experimental results rather than theory. The most commonly used methods for tuning the hyperparameters are grid-search (Chicco, 2017). In this study, this method was applied to find the best possible hyperparameters for each of the weather detection models using the “caret” package in R. A 5-fold cross-validation was used to evaluate the performance of the models. For the SVM models, two parameters (gamma and cost) can be tuned. As shown in Figure 12a and Figure 12b, the best performance was achieved for gamma and cost values of 0.001 and 20, respectively, for the SVM model using HOG; and 0.01 and 20, respectively, for the SVM model using LBP. Note that several kernel functions, including linear, polynomial, radical, and exponential, were considered during the tuning process. However, the radial kernel function produced the best performance.

Considering the RF model, two parameters can be tuned: the number of trees to grow (ntree) and the number of variables randomly sampled at each tree node (mtry). According to Figure 12c and Figure 12d, the best combination for the “ntree” and “mtry” was found to be 600 and 20, respectively, for the RF model using HOG; and 500 and 20, respectively, for the RF model using LBP. Similarly, during the training of the GB model, two parameters, including the number of trees (ntree) and the number of splits on a tree (interaction depth), were tuned. The best detection accuracy was found for an “interaction depth” of 12 and an “ntree” of 700 for the GB model using HOG, as shown in Figure 12e. Also, the best performance was found for an “interaction depth” of 12 and an “ntree” of 600 for the GB model using LBP, as seen in Figure 12f. Note that for the multilevel model, hyperparameters were also tuned for each of the models inside the hierarchical structure.

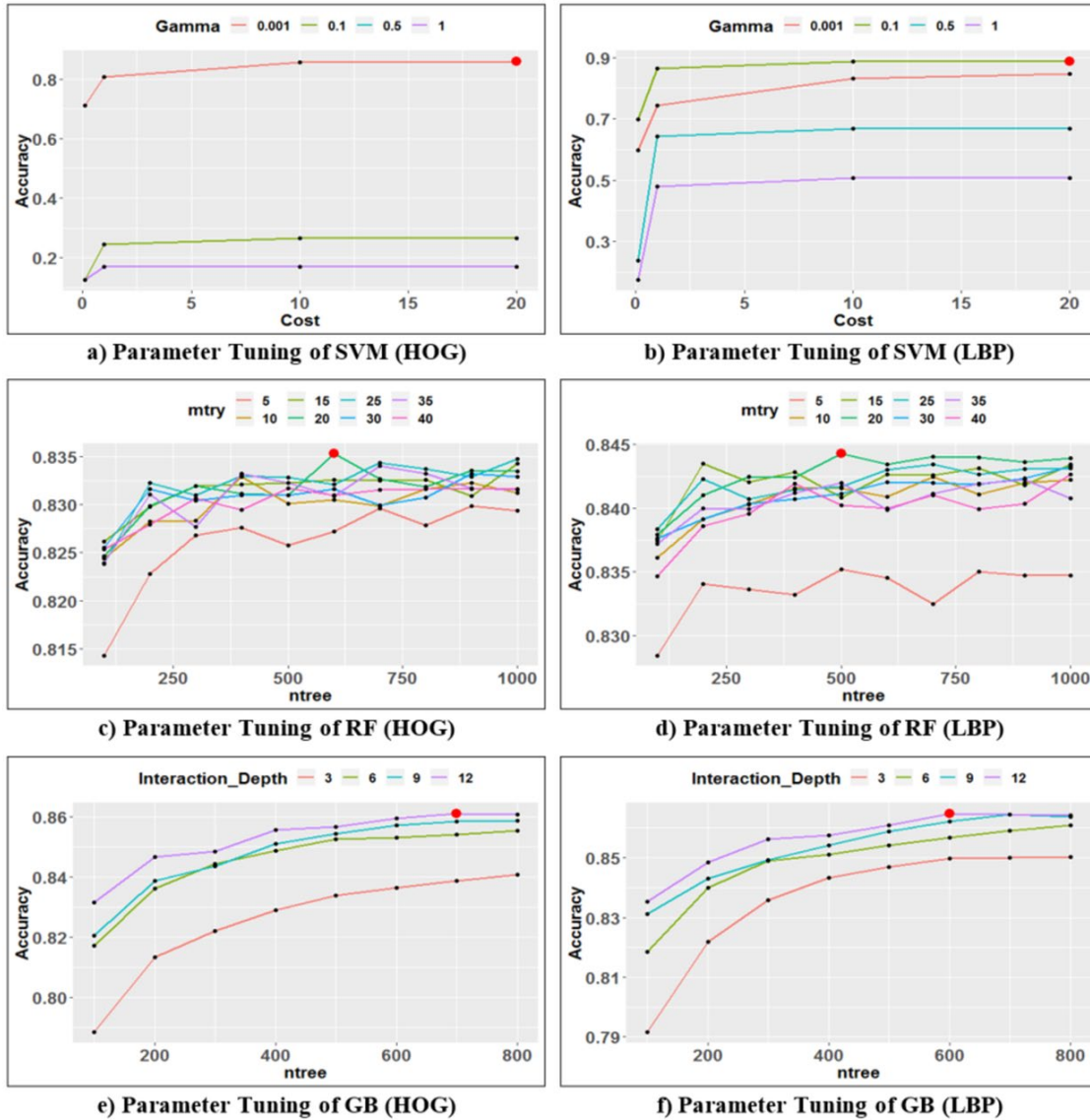


Figure 12. Parameter Tuning of the Weather Detection Model for Image Dataset-1

5.1.3 Performance of the weather detection model using HOG feature

The detection accuracy of the SVM models was found to be around 86%, meaning 86% of the images were classified correctly. Also, the near fog image group had the highest true positive (TP) rate and the lowest false negative (FN) rate, where only about 94% of the near fog images were correctly classified. The lowest accuracy was found for the light rain image group with an accuracy of about 78%. The reason for this low accuracy is because light rain is very similar to light snow and clear weather. However, for extremely adverse weather conditions, e.g., heavy rain, light rain, and near fog, the detection accuracy was found to be more than 90% as seen in Table 7.

The overall detection accuracy of the trained RF model was found to be around 82%. The near fog image group had the highest TP rate with 90% correct classification, whereas the light rain image group had the lowest TP rate with 29% misclassification. Similarly, the prediction accuracy of the heavy rain, light rain, light snow, heavy snow, and distant fog was found to be around 72%, 86%, 80%, 87%, and 76%, respectively, using the trained RF model.

The trained GB model performed slightly worse than the RF model, with an overall prediction accuracy of around 81%. Again, the near fog image group had the highest accuracy with only 10% misclassification; more precisely, 3.4 %, 0.4%, 1%, 4.6%, 0.2%, and 0.4% of the near fog images were classified as distant fog, heavy snow, light snow, heavy rain, light rain, and clear, respectively.

The most promising results were found for the multilevel weather detection model. This modeling technique produced significantly better detection accuracy compared with the other models due to several reasons. First, the texture image features of the image groups might have some similarities. For instance, the light rain and the light snow image groups might have some similar properties. Therefore, if a single model is used for detecting all the weather conditions, there will be a greater possibility of getting misclassification, especially for categories with similar features. Since the proposed multilevel structure detected the levels of adverse weather using separate independent models, it reduced such misclassification. Secondly, a particular classifier may produce better prediction accuracy for a particular type of weather condition. For example, SVM might produce better detection accuracy for identifying levels of rain, e.g., light rain and heavy rain, but might not necessarily perform well for detecting levels of fog, e.g., distant fog and near fog. Since the proposed multilevel structure used four independent models, each being the best performing model for that particular weather group, it improved the overall detection accuracy. The overall detection accuracy was found to be around 89%, which is around 3% higher than the SVM model and around 8% higher than the RF and GB model. The detection accuracy for most of the image categories was found to be more than 90%, as seen in Table 7. The highest accuracy (95%) was found for the near fog image group, and the lowest accuracy (80%) was found for the light rain image group.

Table 7. Performance of the Weather Detection Models Using HOG Features

Model	Performance Measure	Clear	Light Rain	Heavy Rain	Light Snow	Heavy Snow	Distant Fog	Near Fog
SVM	TP Rate/Sensitivity	84.2	78.2	90.4	84.4	90.6	79.8	94.2
	FN Rate	15.8	21.8	9.6	15.6	9.4	20.2	5.8
	Overall Accuracy	86.0						
RF	TP Rate/Sensitivity	81.0	71.6	86.2	80.0	87.0	76.0	90.0
	FN Rate	19.0	28.4	13.8	20.0	13.0	24.0	10.0
	Overall Accuracy	81.7						
GB	TP Rate/Sensitivity	80.8	71.4	85.8	80.6	86.7	74.2	90.0
	FP Rate	2.8	5.1	3.2	3.4	2.0	3.8	1.4
	Overall Accuracy	81.3						
Multilevel Model	TP Rate/Sensitivity	92.4	80.2	92.6	90.8	93.8	82.2	95.0
	FN Rate	7.6	19.8	7.4	9.2	6.2	17.8	5.0
	Overall Accuracy	89.2						

5.1.4 Performance of the weather detection model using LBP feature

The detection accuracy of the weather detection models utilizing the LBP feature is listed in Table 8. Overall, the performance of the models is better than the models calibrated using HOG features. The overall detection accuracy of the model using SVM was around 89%; specifically, the clear, light rain, heavy rain, light snow, heavy snow, distant fog, and near fog images had a detection accuracy of 91%, 83%, 91%, 91%, 91%, 87%, and 92% respectively.

The trained RF model using the LBP feature provided an overall detection accuracy of 82.4%, which is marginally higher than the RF model using the HOG feature. The lowest FN rate was found for near fog, with only 10% misclassification, whereas the highest FN rate was found for light rain with 23% misclassification.

Considering the GB model with LBP features, the overall detection accuracy was 82%, which is slightly lower than the detection accuracy of the RF model, as seen in Table 8. The highest TP rate was found for the near fog image group, and the lowest TP rate was found for the light rain image group.

As expected, the multilevel weather detection model performed better compared with the other models with an impressive overall accuracy of about 91%; specifically, the clear, light rain, heavy rain, light snow, heavy snow, distant fog, and near fog images had a detection accuracy of about 95%, 84%, 92%, 90%, 93%, 88%, and 94%, respectively. The false positive (FP) rate of the clear image group was found to be only 1.6%, which means only 1.6% of other adverse weather conditions were classified as clear weather. Note that from a safety perspective, the higher the FP rate, especially in clear weather, the greater the associated risk because in such scenarios, drivers will be exposed to adverse weather without any prior warning. In addition, a high FN rate of clear weather will create frequent false alarms because it will give drivers adverse weather warnings in clear roadway conditions. This could grow disrespect for the system and might increase compliance issues. The FN rate for the multilevel weather detection model based on LBP features was only 5%, as seen in Table 8.

Table 8. Performance of the Weather Detection Models Using LBP Features

Model	Performance Measure	Clear	Light Rain	Heavy Rain	Light Snow	Heavy Snow	Distant Fog	Near Fog
SVM	TP Rate/Sensitivity	91.4	82.8	91.0	91.0	91.0	87.0	92.4
	FN Rate	8.6	17.2	9.0	9.0	9.0	13.0	7.6
	Overall Accuracy	89.4						
RF	TP Rate/Sensitivity	86.0	77.4	82.0	78.6	82.0	80.8	90.0
	FN Rate	14.0	22.6	18.0	21.4	18.0	19.2	10.0
	Overall Accuracy	82.4						
GB	TP Rate/Sensitivity	85.8	75.6	82.0	79.0	82.2	79.8	89.8
	FN Rate	14.2	24.4	18.0	21.0	17.8	20.2	10.2
	Overall Accuracy	82.0						
Multilevel Model	TP Rate/Sensitivity	95.0	83.6	92.0	90.2	93.0	88.4	94.2
	FN Rate	5.0	16.4	8.0	9.8	7.0	11.6	5.8
	Overall Accuracy	91.0						

5.1.5 Comparison of the computational cost of the weather detection models

The extraction time of the image features based on LBP and HOG, as well as the training time of the weather detection models, are provided in Table 9. All the computations were performed on a workstation with an Intel® Xeon CPU E3-1240 3.4 GHz processor, 24 GB of RAM, and an NVIDIA® Quadro K620 GPU. The feature extraction times of image dataset-1, which consisted of 17,500 images, were 82.3 seconds and 114.4 seconds for HOG and LBP-based features, respectively. As mentioned earlier, this dataset was used to train, validate, and test the SVM, RF, and GBT models. However, the total feature extraction time of the multilevel model was found to be 180.2 and 240.7 seconds for HOG and LBP-based features, respectively. The multilevel model required a higher extraction time since it leveraged four image datasets with 35,000 total images to train four independent models in the hierarchical structure of the multilevel model.

The lowest training times for both HOG and LBP-based models were found for the SVM model followed by the multilevel model, RF and GBM, as shown in Table 9. The HOG-based weather detection models required more time to train compared with LBP-based models since HOG had 240 features compared with only 59 features for LBP. After training, each classifier can instantaneously detect weather conditions from images; therefore, all the trained models can be used in real time to detect weather conditions. Although the computational cost of the multilevel model is relatively higher, its detection accuracy is significantly superior compared with the other detection models. Therefore, considering the available resources and the trade-off between performance and computational cost, transportation practitioners could decide on the selection of the appropriate weather detection model.

Table 9. Computational Cost of the Weather Detection Models

Feature Extraction Method	Model	Dataset Used	Total Number of Images	Feature Extraction Time (s)	Training Time (s)	Overall Accuracy
HOG (240 Features)	SVM	ID-1	17,500	82.3	85.4	86.0
	RF	ID-1	17,500	82.3	194.4	81.7
	GBM	ID-1	17,500	82.3	218.6	81.3
	Multilevel Model	ID-2, ID-3, ID-4, ID-5	35,000	180.2	138.6	89.2
LBP (59 Features)	SVM	ID-1	17,500	114.4	17.4	89.4
	RF	ID-1	17,500	114.4	41.2	82.4
	GBM	ID-1	17,500	114.4	41.4	82.0
	Multilevel Model	ID-2, ID-3, ID-4, ID-5	35,000	240.7	53.0	91.0

*ID = Image Dataset; SVM = Support Vector Machine, RF = Random Forest, GBM = Gradient Boosting Tree

5.1.6 Effect of the number of features on multilevel weather detection

In an attempt to investigate the effect of the number of training features on model performance, a sensitivity analysis was conducted, where the performance of the multilevel model was tested with the increasing number of training features. The number of features using HOG was increased by changing the cell (C) and bin (B) sizes. The cell size determines the amount of spatial information to be captured; therefore, to capture large-scale spatial information, a larger cell size is needed (MathWorks, 2020a). On the other hand, the bin size determines the number of orientation details. Therefore, to capture finer details, a larger bin size is required. By changing these two parameters, the number of features for training the weather detection models using HOG increased from 240 to 980. Similarly, the features of the LBP were increased by changing the number of neighbors (P) around the center pixels. Increasing the value of P provides greater details around each pixel (MathWorks, 2020b). In this analysis, six values of P were used ranging from 8 to 24. The number of features increased from 59 for a P value of 8 to 555 for a P value of 24. As expected, the classification accuracy of the multilevel weather detection model was improved with the increasing number of features at the expense of more computational power, as shown in Table 10. Considering the HOG features, the classification accuracy was improved from 89.2% using 240 features to 95.4% using 980 features. After that, no significant improvement in accuracy was observed. However, the features extraction time and the training time increased from 180.2 seconds to 391.3 seconds and 138.6 seconds to 712.9 seconds, respectively. A similar trend was also observed for the weather detection model based on LBP features. The detection accuracy was improved from 91% using 59 features to 96.4% using 555 features. However, as expected, both the feature extraction time and the training time increased significantly, as listed in Table 10. Therefore, keeping the practical aspects in mind, such as applications in a CV environment, this study suggests the use of a lower number of features when the weather detection model needs to be trained and run on a smartphone platform due to its fewer requirements for computational resources. However, for other cases, when the weather detection model could be trained and applied off-road, such as in the Traffic Management Center (TMC), utilization of more features is suggested due to its relatively higher detection performance.

Table 10. Effect of the Number of Features on Multilevel Weather Detection

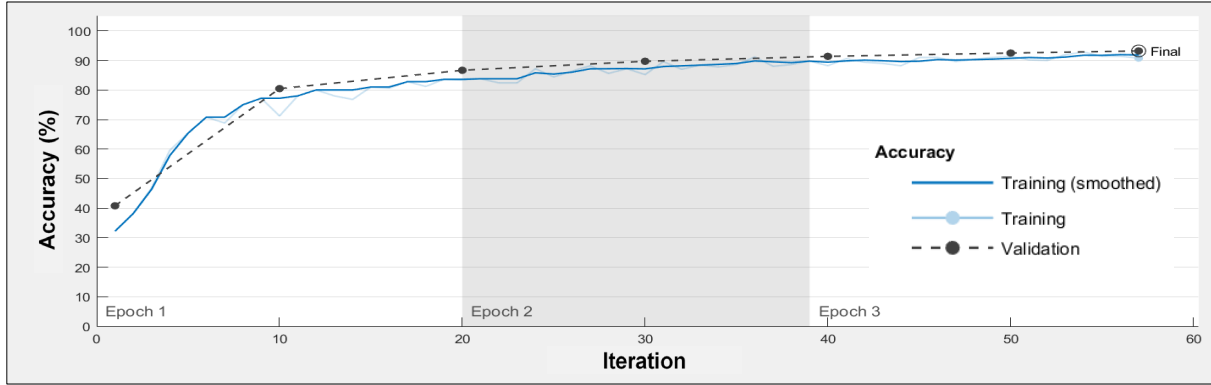
Feature Extraction Method	Configuration	Number of features extracted	Feature Extraction Time (s)	Training Time (s)	Overall Accuracy (%)
HOG	$C = 50 \times 50, B = 5$	240	180.2	138.6	89.2
	$C = 50 \times 50, B = 7$	336	228.7	265.8	91.9
	$C = 40 \times 40, B = 5$	400	234.8	336.1	93.4
	$C = 40 \times 40, B = 7$	560	263.3	494.3	95.1
	$C = 30 \times 30, B = 5$	700	333.0	543.3	94.5
	$C = 30 \times 30, B = 7$	980	391.3	712.9	95.4
LBP	$P = 8$	59	240.7	53.0	91.0
	$P = 12$	135	398.9	86.9	93.1
	$P = 16$	243	420.4	131.1	95.4
	$P = 20$	383	591.5	212.4	96.2
	$P = 22$	465	709.7	276.13	96.2
	$P = 24$	555	748.3	432.1	96.4

* C = Cell size, B = Bin size, P = Number of neighbors

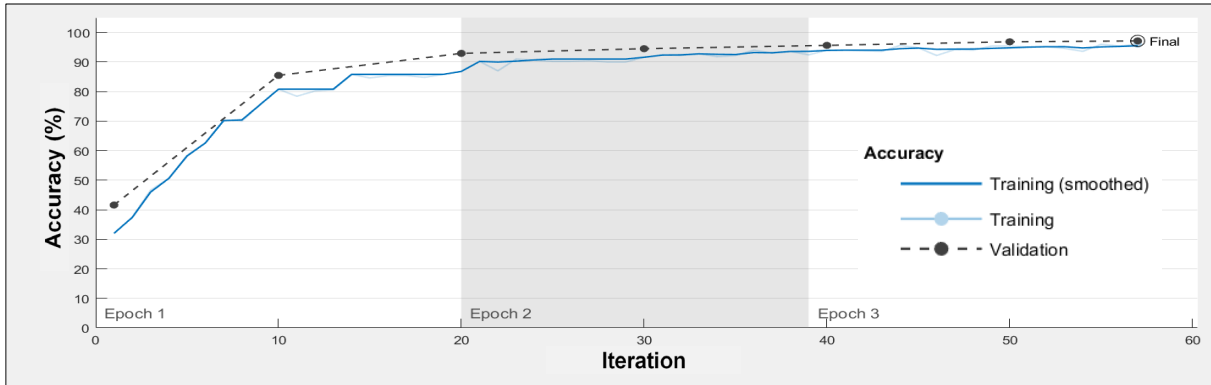
5.2 Weather and Surface Condition Detection from Webcams

5.2.1 Training and validation

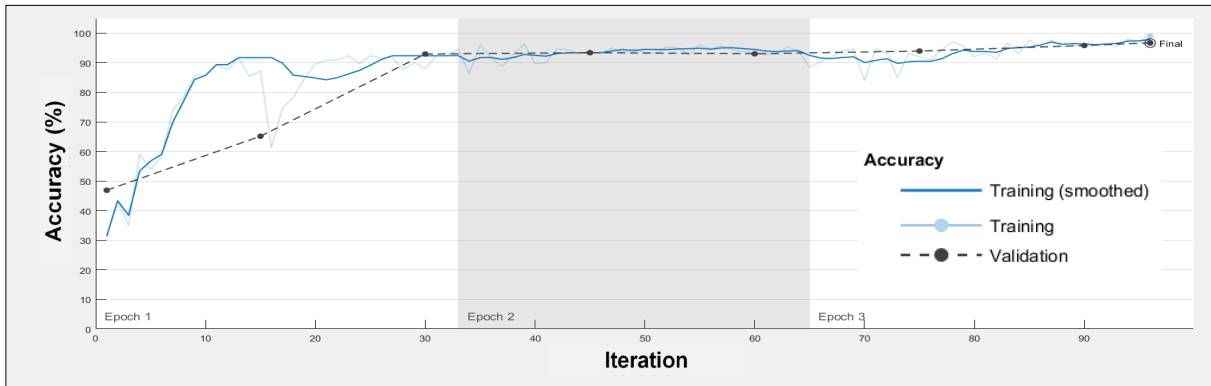
The training progress of the weather detection models along with training and validation accuracy is illustrated in Figure 13. Note that for validation, the holdout technique has been applied where 80% (a total of 9,600 images with 3,200 images in each category) of the training and validation data were used to train the weather detection models, and the remaining 20% (a total of 2,400 images with 800 images in each category) were used for validation. The ResNet18 was found to be the best performing model during validation with an overall accuracy of around 97% at the final iteration. After several trials, the best set of parameters for this model was found to be as follows: maximum epochs = 3, batch size = 300, optimizer: stochastic gradient descent (SGD), initial learning rate = 0.001, learning rate drop factor = 0.1, and learning rate drop period = 8. The GoogLeNet also provided nearly identical performance, the accuracy of which gradually increased from 31% at the first iteration to 97% at the final iteration. Similarly, the AlexNet also produced a very good validation accuracy of 93% at the final iteration after three epochs of training.



a



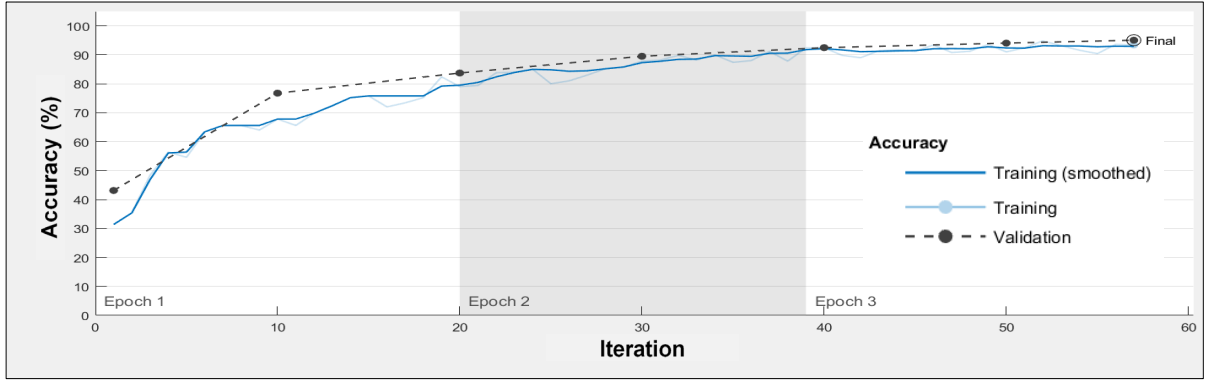
b



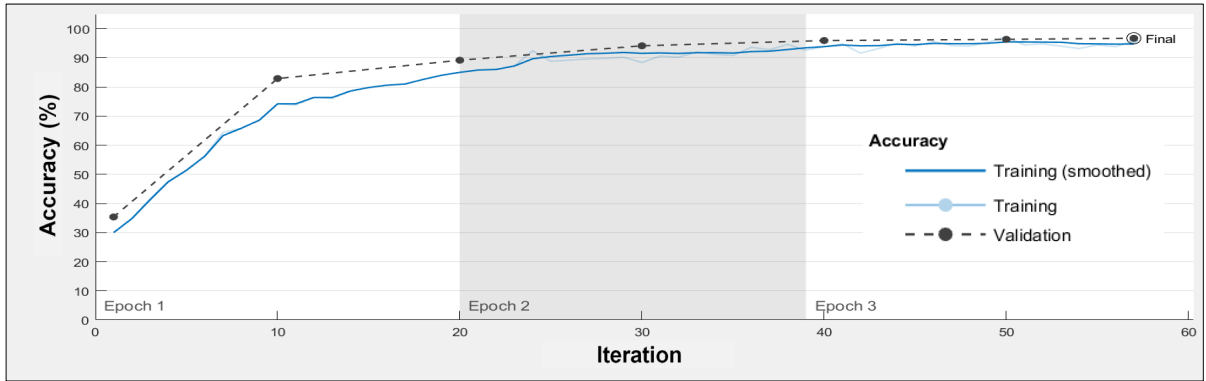
c

Figure 13. Training Progress of Weather Detection Models, a) AlexNet, b) GoogLeNet, c) ResNet18

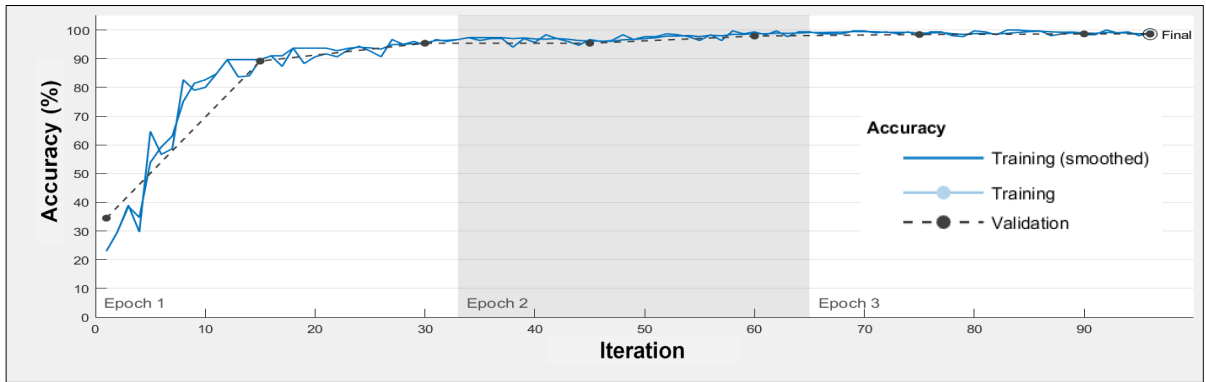
The training progress of the surface condition detection models along with training and validation accuracy is illustrated in Figure 14, which shows that the ResNet18 model provided the best performance during validation. To optimize all the surface condition detection models, SGD with an initial learning rate of 0.0001 was used at every iteration. Instead of using a constant learning rate for parameter updates, SGD uses adaptive learning rates for each parameter, which enables it to provide the best optimization within the least possible time (Khan and Ahmed, 2020). The overall validation accuracy of the ResNet18 model was about 65% at the first iteration, which gradually improved and reached around 98% at the final iteration after three epochs of training. Conversely, AlexNet and GoogLeNet produced an overall validation accuracy of about 95% and 97%, respectively, at the end of the training.



a



b



c

Figure 14. Training Progress of Surface Condition Detection Models, a) AlexNet, b) GoogLeNet, c) ResNet18

In an attempt to further investigate the performance of the pre-trained CNN, models K-fold cross-validation technique was also applied. K-fold cross-validation is a model validation technique that randomly splits the dataset into K number of sets or folds of approximately equal size. Subsequently, each K fold is considered as validation data and the remaining $K - 1$ folds are used to train the machine learning model. This study used 5-fold cross-validation because a value of $K = 5$ is recommended by many researchers and is widely used in the field of applied machine learning (James et al., 2017). The training and validation dataset used in this study for weather detection consisted of 12,000 images with 4,000 images in each weather category. Therefore, each fold consisted of 2,400 images with 800 images in each weather category. The validation and training dataset for surface condition detection also had

similar kinds of splits. The accuracy using 5-fold cross-validation along with holdout validation are listed in Table 11. Interestingly, 5-fold cross-validation and holdout validation provided similar performance in terms of accuracy. The highest overall cross-validation accuracy of 97.4% for detecting weather conditions was found for the ResNet18 model. The GoogLeNet also produced a marginally lower accuracy of 96.9%. Considering surface condition detection, again the ResNet18 produced the best performance with an overall cross-validation accuracy of 97.6%.

Table 11. Model Performance During Validation

	Model	5-fold Cross-Validation						Holdout Validation Accuracy
		Fold 1 Accuracy	Fold 2 Accuracy	Fold 3 Accuracy	Fold 4 Accuracy	Fold 5 Accuracy	Overall Accuracy	
Weather Detection	AlexNet	92.9	93.0	93.7	94.2	94.8	93.7	93.2
	GoogLeNet	95.9	97.1	96.1	97.4	97.9	96.9	96.6
	ResNet18	97.4	97.0	96.4	97.8	98.4	97.4	97.1
Surface Condition Detection	AlexNet	94.9	95.4	96.2	96.1	96.6	95.9	95.1
	GoogLeNet	96.6	95.9	97.6	98.4	98.1	97.3	96.7
	ResNet18	97.1	96.2	98.3	97.8	98.7	97.6	98.5

5.2.2 Performance evaluation

After the training and validation, the surface and weather detection models' performance was evaluated using a test dataset consisting of 20% of the original images, which corresponds to a total of 3,000 images with 1,000 images in each weather category. The quality of the models was evaluated using seven performance indices: overall accuracy, precision, sensitivity, specificity, F1-score, FPR, and FNR. All the pre-trained models provided a high degree of performance with an accuracy of more than 92%, as seen in Table 12. Starting with the overall accuracy for the surface condition detection models, the ResNet18 provided the best performance with an unprecedented overall detection accuracy of 99.1%, which is in accordance with the accuracy (98%) found during validation. Other performance indices were also found to be higher for the ResNet18 model compared with the other two models. However, the training time of this model was 106 minutes on a single CPU, which was significantly higher than the AlexNet. However, the AlexNet provided the least performance with an overall detection accuracy of 94.7%. This study recommends the use of ResNet18 for surface condition detection considering its superior performance; however, for practical application purposes, transportation practitioners should consider the trade-off between accuracy and computational requirements. If the model needs to train and run on a smartphone platform with relatively less processing power, the AlexNet could be adopted since its simple structure significantly reduces the need for computational power.

All the weather detection models also provided very good results with ResNet18 being the best, where an overall detection accuracy of 97.3% was achieved for correctly detecting three weather classes: clear, light snow, and heavy snow, as seen in Table 12. Other performance measures of this model were also found to be superior compared with the other models. The GooLeNet produced nearly identical performance compared to the ResNet18 but required significantly higher training time. The lowest performance was observed for the AlexNet with an overall detection accuracy of 92.4%. Therefore, based on the performance measures, similar to the surface condition detection models, this study also suggests the use of ResNet18 to develop weather detection models from webcams.

The computational requirements of the proposed weather and surface condition detection models are listed in Table 12. All the computations were performed on a workstation with Intel Xeon CPU @ 3.4 GHz processor, 12 Gb of memory, NVIDIA Quadro K620 graphics, and 64 bits Windows 10 operating system. For both weather and surface condition detection, AlexNet required the least and GoogLeNet required the highest amount of time. Relative training time was also observed by calculating the ratio of other pre-trained models (e.g., GoogLeNet and ResNet18) with the AlexNet, which shows that GoogLeNet and ResNet18 required 5.2 and 2.6 times more computational resources, respectively, compared with AlexNet for weather detection. Similar results were also observed for surface condition detection. Note that after training, each model can instantaneously detect weather and surface conditions from images; therefore, all the models can be used in real time.

The webcams used in this study have a frame rate of 30 fps. However, during the implementation phase, the research team will not consider images from every frame. To reduce the computational requirements, 10 images per minute will be extracted and weather and surface conditions will be updated every five minutes on the TMC website and 511 apps. Also, if multiple conditions are present within these five minutes, a weighted average will be used to get more accurate real-time weather and surface conditions.

Table 12. Performance Measures of The Pre-Trained Weather and Surface Condition Detection Models

	Model	Accuracy (%)	Precision (%)	Recall (%)	Specificity (%)	F1-score (%)	FPR (%)	FNR (%)	Training Time (mins)	Relative Training Time
Weather Detection	AlexNet	92.4	92.5	92.4	96.1	92.4	3.9	7.6	39	1
	GoogLeNet	97.2	97.2	97.2	98.6	97.2	1.4	2.8	203	5.2
	ResNet18	97.3	97.3	97.3	98.6	97.3	1.4	2.7	102	2.6
Surface Condition Detection	AlexNet	94.7	94.7	94.7	97.3	94.7	2.7	5.3	42	1
	GoogLeNet	96.7	96.9	97.0	98.4	97.0	1.6	3	212	5.0
	ResNet18	99.1	99.2	99.6	99.2	99.1	0.8	0.4	106	2.5

In order to obtain more insights from the best performing weather and surface detection models (e.g., models based on ResNet18 architecture), performance measures were calculated for each class, as listed in Table 13, and visualized using confusion matrices, as illustrated in Figure 15. The numbers along the diagonal represent the correct classification of the respective image group and are marked as blue. Considering the surface condition detection model, the highest precision, as well as the highest recall, was found for the dry image group; specifically, out of 1,000 test dry surface images only one image was misclassified as snowy surface and four images were misclassified as wet/slushy surface. The lowest FNR (0.5%) and the lowest FPR (0.3%) were also found for the dry images group. Considering the safety-related practical applications, a high degree of FPR of the dry surface condition is particularly hazardous because it would increase the risk by exposing drivers to affected road surface conditions without warnings. Conversely, a high degree of FNR of dry surface conditions would provide frequent false warnings, which might lead to disrespect for the warning systems and might decrease the compliance rate. The lowest performance was found for the wet/slushy image group with a recall value of 99.5%, where 12 images were misclassified as snowy surface images, as seen from Figure 15a. This could be because slush and snow might look similar, especially, when the snow just starts to melt after a heavy snowfall. Interestingly, out of 1,000 test images, only two slushy surface images and three snowy surface images were misclassified as dry surfaces, which indicates the model's unprecedented ability to detect adverse surface conditions.

Considering the weather detection model, the best performance was observed for the clear image group with a recall value of 99.6%, followed by heavy snow and light snow image groups with recall values of 99.2% and 99.3%, respectively, as seen in Table 13. Out of 1,000 clear images only two and five images were wrongly classified as light snow and heavy snow, respectively, as seen in Figure 15b. The FPR of the clear group was also found to be very low with a value of only 0.8%. In safety applications, a low FPR of clear weather ensures better safety, whereas a low FNR promotes better compliance. The lowest detection was observed for the light snow image group with a recall value of 93.6%; specifically, 51 and 13 light snow images were misclassified as heavy snow and clear images, respectively.

Table 13. Detection Summary of the Trained Weather and Surface Condition Detection Model Using ResNet18

Model	Image Category	Precision (%)	Recall (%)	Specificity (%)	F1-score (%)	FPR (%)	FNR (%)	Overall Accuracy (%)
Weather Detection	Clear	98.5	99.3	99.2	98.9	0.8	0.7	97.3
	Heavy Snow	94.8	99.0	97.3	96.9	2.7	1.0	
	Light Snow	98.6	93.6	99.3	96.0	0.7	6.4	
Surface Condition Detection	Dry	99.5	99.5	99.7	99.5	0.3	0.5	99.1
	Snowy	98.7	99.4	99.3	99.0	0.7	0.6	
	Wet/Slushy	99.3	98.6	99.6	98.9	0.4	1.4	

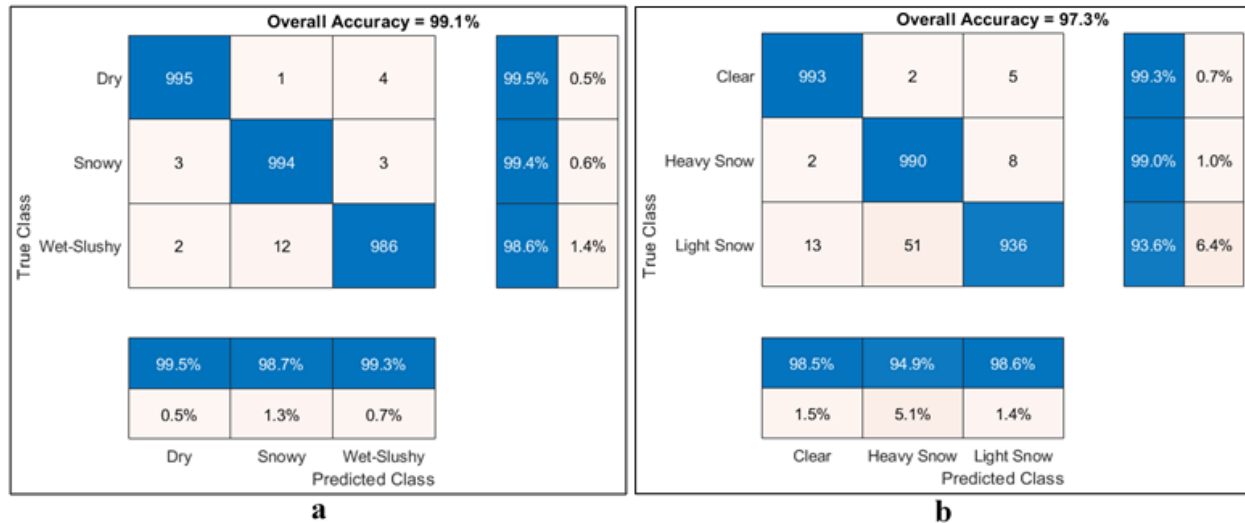


Figure 15. Confusion Matrix of the Trained Models Using Resnet18, A) Surface Condition Detection Model, B) Weather Detection Model

As mentioned earlier, the weather and surface condition detection models were trained using images extracted from a 402-mile corridor of I-80 in Wyoming. However, to evaluate the performance of the proposed detection models on other roadways, a separate image dataset was created using webcam images from various locations on I-25, I-90, US14, US 20, and US 287 in Wyoming. This new image dataset consisted of 300 weather images with 100 images in each weather category and 300 surface images with 100 images in each surface category. Subsequently, the new images were tested using the

previously trained weather and surface condition detection models based on ResNet18 architecture. It was found that 93.7% of the weather images (281 out of 300 images) and 95.3% (286 out of 300 images) of the surface images were correctly classified, indicating that the proposed detection models can perform equally well on other roadways.

5.2.3 Comparison of the proposed pre-trained CNN with traditional machine learning models

In order to compare the results of the proposed weather and surface condition detection models based on pre-trained CNN architecture with the traditional machine learning models, six models, classification and regression tree (CART), k-nearest neighbors (K-NN), support vector machine (SVM), random forest (RF), gradient boosting (GB), and naïve bayes (NB), were trained and comprehensively evaluated using the same image datasets. Since traditional machine learning requires the extraction of image features to train the models, this study leveraged a widely used and effective feature extraction method, local binary pattern (LBP), to extract the image features. LBP compares each pixel of an image to its neighborhood pixels to express the local variation of image texture (Ojala et al., 1996). Although LBP was created considering a fixed 3×3 window to capture the texture variation, this study leveraged an updated version of LBP with two parameters. The first parameter describes the number of pixels around the center pixel and was denoted by P , and the second parameter defines the radius of the circle and was denoted by R . The LBP features can be defined using the following equations.

$$LBP_{P,R} = \sum_{p=0}^{P-1} s(g_p - g_c) 2^p$$

$$s(x) = \begin{cases} 1, & x \geq 0 \\ 0, & x < 0 \end{cases}$$

Here, P is the number of pixels, R is the radius, and g_p and g_c is the gray-level intensity of the surrounding and center pixels, respectively. Following a previous study, this study considered 8 pixels ($P = 8$) around the center pixel and a radius of 1 ($R = 1$), which produced 59 features (Khan and Ahmed, 2019). Subsequently, the extracted image features were used to train, validate, test, and compare the performance of the traditional machine learning models. The comparison revealed that the performance of the pre-trained CNN models, especially GoogLeNet and ResNet18, is superior to the traditional machine learning models in terms of several performance criteria, such as precision, recall, specificity, and accuracy, as shown in Table 14.

Considering weather detection, the highest overall accuracy of 95.4% based on traditional machine learning models was found using the K-NN model. SVM, RF, and GB also produced good detection performance with more than 90% accuracy. However, none of the traditional machine learning models exceeded the detection performance of GoogLeNet and ResNet18. As already mentioned, the highest overall accuracy of 97.3% in detecting weather was achieved with ResNet18.

Surface condition detection models with traditional machine learning also provided a similar trend, as seen in Table 14. The best performance was achieved using the K-NN model with an overall detection accuracy of 92.6%. RF models also produced a nearly identical performance with an overall detection accuracy of 92.4%. However, all the pre-trained CNN models outperformed the traditional machine learning models. As expected, ResNet18 produced the best surface condition detection accuracy with an unprecedented overall accuracy of 99.1%.

Table 14. Comparison of the Pre-Trained CNNs with the Traditional Machine Learning Models

	Models	Precision (%)	Recall (%)	Specificity (%)	Accuracy (%)
Weather Detection	CART	86.2	86.2	93.1	86.1
	K-NN	95.5	95.4	97.7	95.4
	SVM	93.0	93.0	96.5	93.0
	RF	94.9	94.9	97.4	94.9
	GB	90.8	90.8	95.4	90.7
	NB	68.4	68.1	84.0	67.9
	AlexNet	92.5	92.4	96.1	92.4
	GoogLeNet	97.2	97.2	98.6	97.2
	ResNet18	97.3	97.3	98.6	97.3
Surface Condition Detection	CART	77.9	77.9	88.9	77.8
	K-NN	92.6	92.7	96.3	92.6
	SVM	90.4	90.5	95.2	90.4
	RF	92.5	92.4	96.2	92.4
	GB	85.5	85.5	92.7	85.5
	NB	59.3	58.0	79.2	58.4
	AlexNet	94.7	94.7	97.3	94.7
	GoogLeNet	96.9	97.0	98.4	96.7
	ResNet18	99.2	99.6	99.2	99.1

6. CONCLUSIONS AND RECOMMENDATIONS

This research first developed several affordable in-vehicle detection systems that can provide accurate trajectory-level weather information at road surface level with unprecedented accuracy. More than 30,000 images were extracted from the NDS videos in different weather and roadway conditions. Subsequently, the images were categorized into seven weather levels: clear, light rain, heavy rain, light snow, heavy snow, distant fog, and near fog. To train the weather detection models, two features were extracted from the images: HOG and LBP. This study first considered three machine learning algorithms: SVM, RF, and GB. Subsequently, a unique multilevel detection model was also devised, where four machine learning algorithms were merged in a hierarchical structure to achieve better detection accuracy. In addition, instead of using the default values, the hyperparameters of all models were tuned by carefully observing the training process using 5-fold cross-validation. Finally, the performance of the weather detection model was evaluated using a separate test dataset. The overall detection accuracies of the SVM, RF, and GB model using HOG features were found to be around 86%, 82%, and 81%, respectively. However, the multilevel detection models outperformed the overall models with an overall detection accuracy of 89%. The detection models using the LBP feature performed slightly better, where SVM, RF, and GB produced an overall accuracy of 89%, 82%, and 82%, respectively. For the LBP feature-based models, the multilevel approach was also found to be superior compared with the other models with an impressive overall accuracy of 91%.

The research then concentrated on the transferability of the findings in the state of Wyoming and successfully developed Wyoming-specific weather and surface condition detection systems. The study utilized transfer learning techniques to train the detection models based on three pre-trained convolution neural network (CNN) architectures: AlexNet, GoogLeNet, and ResNet18. The last few layers of the pre-trained models were carefully modified to meet the classification need. Two annotated image datasets were created from the webcam images: one for surface condition detection with three categories, including dry, snowy, and wet/slushy; and the other one for weather detection also with three categories, including clear, light snow, and heavy snow. For each of these detection models, 80% of the data were used during training and validation and the rest were used to test the detection quality of the trained models based on several performance indices, such as precision, recall, specificity, F1-score, false negative rate, and false positive rate. The ResNet18 architecture produced the best performance in terms of these indices with impressive overall weather and surface condition detection accuracy of around 97% and 99%, respectively.

The results of this research can be effectively used to improve the weather-based VSL systems in a connected vehicle (CV) environment. Most of the current weather-based regulatory VSL systems are mainly based on the data collected from road weather information system (RIWS) stations. However, weather stations are costly, and hence, their widespread implementation is not feasible. Since the proposed method only uses a single video camera to detect weather, for non-VSL corridors, it has the potential to provide low-cost advisory VSL. To maintain the homogeneity of the speed limit, the weather data from all the available vehicles within a VSL corridor will be considered and only one representative weather data will be used to calculate the speed limit at the beginning of the corridor. In addition, the system can also be used to disseminate cautionary messages within the advanced traveler information system (ATIS), such as “Dense Fog 1-mile Ahead” over the DMS to warn the drivers about any potentially hazardous weather in roadways where no RWIS is present. Another potential application of this study is to develop an affordable advanced driver assistance system (ADAS). The proposed weather detection method is cost-efficient, simple, and more importantly, can be made widely available mainly due to the recent boom of very capable smartphone camera systems. The methodology proposed in this study could also be used to detect work zones, pedestrians, lane changes, motor vehicle crashes, and road closures. In a CV environment, this information can easily be shared with other road users and traffic

management centers (TMCs) to create and disseminate appropriate warnings based on real-time roadway conditions in a smartphone platform.

The proposed models have numerous readily implementable practical applications, most importantly, with no extra cost since the models are developed based on the existing roadside webcams. Note that most DOTs operate and maintain too many roadside webcams for various purposes, including traffic surveillance and manual weather detection. The current practice is to review and process the video feeds from these webcams manually by the TMC operators, which is not always feasible due to time and resource constraints. In addition, processing such huge amounts of information manually and linking it to the corresponding road networks is often subject to human error. The models proposed in this study have the potential to solve these issues by automatically detecting real-time road weather and surface conditions and by linking this information to the corresponding road networks without any human involvement.

Most of the major interstates and highways in Wyoming have fixed webcams with each location having three views of the roadway: upstream, downstream, and road surface. For instance, the 402-mile corridor on I-80 in Wyoming, which is the study area of this research, currently has 56 roadside webcams (WYDOT, 2021). Using the proposed detection models, real-time weather and surface conditions can be automatically detected at each webcam location with excellent detection accuracy without any human involvement. Furthermore, since all the webcams along a road network are usually kept running 24 hours a day, temporal and spatial variations of weather and surface conditions could also be generated using appropriate buffer zones around each webcam location. If any location has multiple webcams, then weighted average could be used to get more accurate real-time weather and surface conditions. Based on this real-time information, cautionary messages, such as “Slippery Roadway Ahead” or “Heavy Snow in Next 5 miles” could be generated and disseminated via DMS to the road users. In addition, using the historical data from these webcams, weather prediction might also be possible. All this useful information can be integrated to build a comprehensive ATIS, which then can be made available to road users via TMC websites, radios, and mobile apps, such as 511. The implementation of the proposed weather and surface condition detection models will be comprehensively explored in future studies.

As mentioned earlier, temporal and spatial variations of weather and surface conditions could be determined utilizing the proposed detection models. This information is highly beneficial for winter maintenance to ensure optimum utilization of available resources. In other words, the developed models could be leveraged to automatically select optimum routes and times for maintenance vehicles such as snowplows. In addition, the proposed detection models could be easily calibrated and integrated using video feeds from snowplows. Recently, many DOTs, including WYDOT, implemented an innovative road condition monitoring system using tablets mounted in snowplows and maintenance vehicles (WYDOT, 2019). The monitoring system requires drivers to report weather conditions manually by tapping nine codes on the tablet touchscreen while driving. This may cause inconsistent reporting of weather conditions because of the variations in driver perception and the subjectivity in reporting different conditions. More importantly, this system may pose some risks to drivers, especially, with their very challenging driving environment during adverse weather conditions. With proper calibration, the proposed models have the potential to detect weather and surface conditions automatically, even from cameras mounted on snowplows, which will eliminate the need for manual reporting.

7. REFERENCES

1. Federal Highway Administration. "How Do Weather Events Impact Roads?" FHWA Road Weather Management. https://ops.fhwa.dot.gov/weather/q1_roadimpact.htm. Accessed Sep. 20, 2021.
2. Perry, A. H., and L. Symons. *Highway Meteorology*. Taylor & Francis Books, Inc, 2003.
3. Andrey, J., B. Mills, M. Leahy, and J. Suggett. "Weather as a Chronic Hazard for Road Transportation in Canadian Cities." *Natural Hazards*, Vol. 28, 2003, pp. 319–343.
4. Ahmed, M., H. Huang, M. Abdel-Aty, and B. Guevara. "Exploring a Bayesian Hierarchical Approach for Developing Safety Performance Functions for a Mountainous Freeway." *Accident Analysis & Prevention*, Vol. 43, No. 4, 2011, pp. 1581–1589. <https://doi.org/10.1016/J.AAP.2011.03.021>.
5. Brown, Harry. "The 20 Largest Traffic Accidents in Recent US History." <https://www.harrybrownlaw.com/blog/20-largest-traffic-accidents-in-united-states>. Accessed Mar. 28, 2018.
6. Moore, R. L., and L. Cooper. *Fog and Road Traffic*. Crowthorne, 1972.
7. Codling, P. J. "Thick Fog and Its Effect on Traffic Flow and Accident." *Transport and Road Research Laboratory, TRRL Report LR 397*, 1971.
8. Brodsky, H., and A. S. Hakkert. "Risk of Road Accident in Rainy Weather." *Accident Analysis and Prevention*, Vol. 20, No. 3, 1988, pp. 161–176.
9. Andreescu, M. P., and D. B. Frost. "Weather and Traffic Accidents in Montreal, Canada." *Climate Research*, Vol. 9, No. 3, 1998, pp. 225–230.
10. WYDOT. Wyoming Report on Traffic Crashes – 2018. *Highway Safety Program, Wyoming Department of Transportation, Cheyenne, WY*, 2019.
11. Saha, Promotes, Mohamed M. Ahmed, and Rhonda Kae Young. "Safety Effectiveness of Variable Speed Limit System in Adverse Weather Conditions on Challenging Roadway Geometry." *Transportation Research Record: Journal of the Transportation Research Board*, No. 2521, 2015, pp. 45–53. <https://doi.org/10.3141/2521-05>.
12. Sui, Y., and R. Young. "Impact of Dynamic Message Signs on Speeds Observed on a Rural Interstate." *Journal of Transportation Engineering*, Vol. 140, No. 6, 2014.
13. WYDOT. Snowplow Priority Plan. *Wyoming Department of Transportation, Cheyenne, WY*, 2019.
14. U.S. Department of Transportation. RITA | ITS | Costs: Road Weather Information System (RWIS) Environmental Sensor Station (ESS). <https://www.itskrs.its.dot.gov/its/benecost.nsf/ID/7f42161d63dbc2c78525852d0050308a>. Accessed Apr. 21, 2020.

15. Wiley, V., and T. Lucas. "Computer Vision and Image Processing: A Paper Review." *International Journal of Artificial Intelligence Research*, Vol. 2, No. 1, 2018, p. 22. <https://doi.org/10.29099/ijair.v2i1.42>.
16. Jonsson, P. "Road Condition Discrimination Using Weather Data and Camera Images." *IEEE Conference on Intelligent Transportation Systems, Proceedings, ITSC*, Vol. 4460318, 2011, pp. 1616–1621.
17. Carrillo, J., M. Crowley, G. Pan, and L. Fu. "Comparison of Deep Learning Models for Determining Road Surface Condition from Roadside Camera Images and Weather Data." 2019.
18. Pan, G., L. Fu, R. Yu, and M. Muresan. "Evaluation of Alternative Pre-Trained Convolutional Neural Networks for Winter Road Surface Condition Monitoring." 2019.
19. Lee, J., B. Hong, Y. Shin, and Y. J. Jang. "Extraction of Weather Information on Road Using CCTV Video." *2016 International Conference on Big Data and Smart Computing, BigComp 2016*, 2016, pp. 529–531.
20. Lee, J., B. Hong, S. Jung, and V. Chang. "Clustering Learning Model of CCTV Image Pattern for Producing Road Hazard Meteorological Information." *Future Generation Computer Systems*, Vol. 86, 2018, pp. 1338–1350. <https://doi.org/10.1016/j.future.2018.03.022>.
21. Babari, R., N. Hautière, É. Dumont, N. Paparoditis, and J. Misener. "Visibility Monitoring Using Conventional Roadside Cameras - Emerging Applications." *Transportation Research Part C: Emerging Technologies*, Vol. 22, 2012, pp. 17–28. <https://doi.org/10.1016/j.trc.2011.11.012>.
22. Ozcan, K., A. Sharma, S. Knickerbocker, J. Merickel, N. Hawkins, and M. Rizzo. "Road Weather Condition Estimation Using Fixed and Mobile Based Cameras." *Advances in Computer Vision. CVC 2019. Advances in Intelligent Systems and Computing*, Vol. 943, 2020, pp. 192–204.
23. Ibrahim, M. R., and J. Haworth. "WeatherNet: Recognising Weather and Visual Conditions from Street-Level Images Using Deep Residual Learning." *International Journal of Geo-Information*, Vol. 8, No. 12, 2019, p. 549.
24. Chu, W. T., X. Y. Zheng, and D. S. Ding. "Image2weather: A Large-Scale Image Dataset for Weather Property Estimation." *IEEE 2nd International Conference on Multimedia Big Data*, 2016, pp. 137–144. <https://doi.org/10.1109/BigMM.2016.9>.
25. Guerra, J. C. V., Z. Khanam, S. Ehsan, R. Stolkin, and K. McDonald-Maier. "Weather Classification: A New Multi-Class Dataset, Data Augmentation Approach and Comprehensive Evaluations of Convolutional Neural Networks." *2018 NASA/ESA Conference on Adaptive Hardware and Systems (AHS)*, 2018, pp. 305–310.
26. Pomerleau, D. "Visibility Estimation from a Moving Vehicle Using the RALPH Vision System." *Proceedings of IEEE Conference on Intelligent Transportation Systems*, 1997, pp. 906–911.
27. Yan, X., Y. Luo, and X. Zheng. "Weather Recognition Based on Images Captured by Vision System in Vehicle." *In International Symposium on Neural Networks*, 2009, pp. 390–398. https://doi.org/10.1007/978-3-642-01513-7_42.

28. Khan, M. N., and M. M. Ahmed. "Snow Detection Using In-Vehicle Video Camera with Texture-Based Image Features Utilizing K-Nearest Neighbor, Support Vector Machine, and Random Forest." *Transportation Research Record: Journal of the Transportation Research Board*, Vol. 2673, No. 8, 2019, pp. 221–232. <https://doi.org/10.1177/0361198119842105>.
29. Khan, M. N., and M. M. Ahmed. "Trajectory-Level Fog Detection Based on in-Vehicle Video Camera with TensorFlow Deep Learning Utilizing SHRP2 Naturalistic Driving Data." *Accident Analysis and Prevention*, Vol. 142, No. 105521, 2020. <https://doi.org/10.1016/j.aap.2020.105521>.
30. Qian, Y., E. J. Almazan, and J. H. Elder. "Evaluating Features and Classifiers for Road Weather Condition Analysis." *International Conference on Image Processing, ICIP*, 2016, pp. 4403–4407. <https://doi.org/10.1109/ICIP.2016.7533192>.
31. Kawai, S., K. Takeuchi, K. Shibata, and Y. Horita. "A Smart Method to Distinguish Road Surface Conditions at Night-Time Using a Car-Mounted Camera." *IEEJ Transactions on Electronics, Information and Systems*, Vol. 134, No. 6, 2014, pp. 878–884. <https://doi.org/10.1541/ieejieiss.134.878>.
32. Bronte, S., L. M. Bergasa, and P. F. Alcantarilla. "Fog Detection System Based on Computer Vision Techniques." *2009 12th International IEEE Conference on Intelligent Transportation Systems*, 2009. <https://doi.org/10.1109/ITSC.2009.5309842>.
33. Gallen, R., A. Cord, N. Hautière, and D. Aubert. "Towards Night Fog Detection through Use of In-Vehicle Multipurpose Cameras." *IEEE Intelligent Vehicles Symposium*, 2011, pp. 399–404. <https://doi.org/10.1109/IVS.2011.5940486>.
34. Hautière, N., R. Labayrade, and D. Aubert. "Real-Time Disparity Contrast Combination for Onboard Estimation of the Visibility Distance." *IEEE Transactions on Intelligent Transportation Systems*, Vol. 7, No. 2, 2006, pp. 201–211.
35. Das, A., M. N. Khan, and M. M. Ahmed. "Detecting Lane Change Maneuvers Using SHRP2 Naturalistic Driving Data: A Comparative Study Machine Learning Techniques." *Accident Analysis and Prevention*, Vol. 142, No. 105578, 2020. <https://doi.org/10.1016/j.aap.2020.105578>.
36. Das, A., A. Ghasemzadeh, and M. M. Ahmed. "Analyzing the Effect of Fog Weather Conditions on Driver Lane-Keeping Performance Using the SHRP2 Naturalistic Driving Study Data." *Journal of Safety Research*, Vol. 68, 2019, pp. 71–80. <https://doi.org/10.1016/j.jsr.2018.12.015>.
37. Khan, M. N., and M. M. Ahmed. "Development of a Novel Convolutional Neural Network Architecture Named RoadweatherNet for Trajectory-Level Weather Detection Using SHRP2 Naturalistic Driving Data." *Transportation Research Record: Journal of the Transportation Research Board*, Vol. 2675, No. 9, 2021, pp. 1016–1030. <https://doi.org/10.1177/03611981211005470>.
38. Das, A., and M. M. Ahmed. "Machine Learning Approach for Predicting Lane Change Maneuvers Using the SHRP2 Naturalistic Driving Study Data." *Transportation Research Record: Journal of the Transportation Research Board*, Vol. 2675, No. 9, 2021, pp. 574–594. <https://doi.org/10.1177/03611981211003581>.
39. Hankey, J. M., Miguel A. Perez, and J. A. McClafferty. *Description of the SHRP2 Naturalistic Database and the Crash, near-Crash, and Baseline Data Sets*. 2016.

40. Smadi, O. SHRP2 - Roadway Information Database. *Iowa State University CTRE*.
41. Ahmed, M. M., A. Ghasemzadeh, B. Hammit, M. N. Khan, A. Das, E. Ali, R. K. Young, and H. Eldeeb. "Driver Performance and Behavior in Adverse Weather Conditions: An Investigation Using the SHRP2 Naturalistic Driving Study Data-Phase 2." *FHWA-WY-18/05F*, 2018.
42. Ahmed, M. M., R. K. Young, A. Ghasemzadeh, B. Hammit, E. Ali, M. N. Khan, A. Das, and H. Eldeeb. "Implementation of SHRP2 Results within the Wyoming Connected Vehicle Variable Speed Limit System: Phase 2 Early Findings Report and Phase 3 Proposal." *U.S. Department of Transportation*, 2017.
43. Khan, M. N., A. Das, and M. M. Ahmed. "Non-Parametric Association Rules Mining and Parametric Ordinal Logistic Regression for an In-Depth Investigation of Driver Speed Selection Behavior in Adverse Weather Using SHRP2 Naturalistic Driving Study Data." *Transportation Research Record: Journal of the Transportation Research Board*, Vol. 2674, No. 11, 2020, pp. 101–119. <https://doi.org/10.1177/0361198120941509>.
44. Stocker, M., P. Silvonen, M. Rönkkö, and M. Kolehmainen. "Detection and Classification of Vehicles by Measurement of Road-Pavement Vibration and by Means of Supervised Machine Learning." *Journal of Intelligent Transportation Systems: Technology, Planning, and Operations*, Vol. 20, No. 2, 2016, pp. 125–137. <https://doi.org/10.1080/15472450.2015.1004063>.
45. Albousefi, A. A., H. Ying, D. Filev, F. Syed, K. O. Prakah-Asante, F. Tseng, and H. Yang. "A Two-Stage-Training Support Vector Machine Approach to Predicting Unintentional Vehicle Lane Departure." *Journal of Intelligent Transportation Systems*, Vol. 21, No. 1, 2017, pp. 41–51. <https://doi.org/10.1080/15472450.2016.1196141>.
46. Gaweesh, S. M., M. N. Khan, and M. M. Ahmed. "Development of a Novel Framework for Hazardous Materials Placard Recognition System to Conduct Commodity Flow Studies Using Artificial Intelligence AlexNet Convolutional Neural Network." *Transportation Research Record: Journal of the Transportation Research Board*, Vol. In-Press, 2021, pp. 1–18.
47. Chern, N. N. K., P. A. Neow, and M. H. Ang. "Practical Issues in Pixel-Based Autofocusing for Machine Vision." *IEEE Explore*, No. 3, 2001, pp. 2791–2796.
48. Dalal, N., and B. Triggs. *Histograms of Oriented Gradients for Human Detection*. 2005.
49. Song, F., X. Tan, X. Liu, and S. Chen. "Eyes Closeness Detection from Still Images with Multi-Scale Histograms of Principal Oriented Gradients." *Pattern Recognition*, Vol. 47, No. 9, 2014, pp. 2825–2838. <https://doi.org/10.1016/j.patcog.2014.03.024>.
50. Tian, S., U. Bhattacharya, S. Lu, B. Su, Q. Wang, X. Wei, Y. Lu, and C. L. Tan. "Multilingual Scene Character Recognition with Co-Occurrence of Histogram of Oriented Gradients." *Pattern Recognition*, Vol. 51, 2016, pp. 125–134. <https://doi.org/10.1016/j.patcog.2015.07.009>.
51. Strang, G. *Introduction to Linear Algebra*. Wellesley-Cambridge Press, Massachusetts, 2009.
52. Suard, F., A. Rakotomamonjy, A. Bensrhair, and A. Broggi. "Pedestrian Detection Using Infrared Images Histograms of Oriented Gradients." *Proceedings of the 2006 IEEE Intelligent Vehicles Symposium*, 2006.

53. Ojala, T., M. Pietikäinen, and D. Harwood. "A Comparative Study of Texture Measures with Classification Based on Featured Distributions." *Pattern Recognition*, Vol. 29, No. 1, 1996, pp. 51–59. [https://doi.org/10.1016/0031-3203\(95\)00067-4](https://doi.org/10.1016/0031-3203(95)00067-4).
54. Ojala, T., M. Pietikäinen, and T. Mäenpää. "Multiresolution Gray-Scale and Rotation Invariant Texture Classification with Local Binary Patterns." *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 24, No. 7, 2002, pp. 971–987. <https://doi.org/10.1109/TPAMI.2002.1017623>.
55. Ahonen, T., A. Hadid, and M. Pietikäinen. "Face Description with Local Binary Patterns: Application to Face Recognition." *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 28, No. 12, 2006, pp. 2037–2041. <https://doi.org/10.1109/TPAMI.2006.244>.
56. Friedman, J., T. Hastie, and R. Tibshirani. *The Elements of Statistical Learning*. Springer, New York, US, 2008.
57. Suthaharan, S. *Machine Learning Models and Algorithms for Big Data Classification*. Springer, Boston, MA, 2016.
58. Breiman, L. E. O. "Random Forests." *Machine Learning*, Vol. 108, No. 271, 2001.
59. James, G., D. Witten, T. Hastie, and R. Tibshirani. *An Introduction to Statistical Learning: With Applications in R*. Springer, New York, 2017.
60. Krizhevsky, A., I. Sutskever, and G. E. Hinton. "ImageNet Classification with Deep Convolutional Neural Networks." *Advances in Neural Information Processing Systems*, 2012, pp. 1097–1105.
61. Simonyan, K., and A. Zisserman. "Very Deep Convolutional Networks for Large-Scale Image Recognition." *arXiv preprint arXiv:1409.1556*, 2015.
62. Sermanet, P., D. Eigen, X. Zhang, M. Mathieu, R. Fergus, and Y. Lecun. "OverFeat: Integrated Recognition, Localization and Detection Using Convolutional Networks." *In ICLR*, 2014.
63. Ke, Q., J. Liu, and M. Bennamoun, S. An, F. Sohel, and F. Boussaid. "Computer Vision for Human-Machine Interaction." *In Computer Vision For Assistive Healthcare*, Elsevier Inc., pp. 127–145.
64. Goodfellow, I., Y. Bengio, and A. Courville. *Deep Learning*. MIT Press, Cambridge, MA, 2016.
65. Nair, V., and G. E. Hinton. "Rectified Linear Units Improve Restricted Boltzmann Machines." *27th International Conference on Machine Learning*.
66. Wani, M. A., F. A. Bhat, S. Afzal, and A. I. Khan. "Basics of Supervised Deep Learning." *Advances in Deep Learning. Studies in Big Data*, Vol. 57, 2020, pp. 13–29.
67. Ardakani, A. A., A. R. Kanafi, U. R. Acharya, N. Khadem, and A. Mohammadi. "Application of Deep Learning Technique to Manage COVID-19 in Routine Clinical Practice Using CT Images: Results of 10 Convolutional Neural Networks." *Computers in Biology and Medicine*, Vol. 121, 2020, p. 103795. <https://doi.org/10.1016/j.combiomed.2020.103795>.

68. Szegedy, C., W. Liu, Y. Jia, P. Sermanet, S. Reed, D. Anguelov, D. Erhan, V. Vanhoucke, and A. Rabinovich. "Going Deeper with Convolutions." *IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 2015, pp. 1–9.
<https://doi.org/10.1109/CVPR.2015.7298594>.
69. He, K., X. Zhang, S. Ren, and J. Sun. "Deep Residual Learning for Image Recognition." *IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, Vol. 2016-Decem, 2016, pp. 770–778. <https://doi.org/10.1109/CVPR.2016.90>.
70. Shi, Y., D. Zhang, J. Wen, X. Tong, X. Ying, and H. Zha. "Radial Lens Distortion Correction by Adding a Weight Layer with Inverted Foveal Models to Convolutional Neural Networks." *24th International Conference on Pattern Recognition*, Vol. 2018-August, 2018, pp. 1894–1899.
<https://doi.org/10.1109/ICPR.2018.8545218>.
71. Suthaharan, S. "Machine Learning Models and Algorithms for Big Data Classification: Thinking with Examples for Effective Learning." *Integrated Series in Information Systems*, Vol. 36, 2016.
72. Kiremire, A. R. "The Application of the Pareto Principle in Software Engineering." *Consult*, Vol. January 13, 2011.
73. Maeda-Gutiérrez, V., C. E. Galván-Tejada, L. A. Zanella-Calzada, J. M. Celaya-Padilla, J. I. Galván-Tejada, H. Gamboa-Rosales, H. Luna-García, R. Magallanes-Quintanar, C. A. Guerrero Méndez, and C. A. Olvera-Olvera. "Comparison of Convolutional Neural Network Architectures for Classification of Tomato Plant Diseases." *Applied Sciences*, Vol. 10, No. 4, 2020.
<https://doi.org/10.3390/app10041245>.
74. Osman, O. A., M. Hajij, P. R. Bakhit, and S. Ishak. "Prediction of Near-Crashes from Observed Vehicle Kinematics Using Machine Learning." *Transportation Research Record: Journal of the Transportation Research Board*, Vol. 2673, No. 12, 2019, pp. 463–473.