2021 NDIA GROUND VEHICLE SYSTEMS ENGINEERING AND TECHNOLOGY SYMPOSIUM
POWER & MOBILITY (P&M) TECHNICAL SESSION
AUGUST 10-12, 2021 - NOVI, MICHIGAN

WINTER TERRAIN CLASSIFICATION TO DETERMINE WHERE A VEHICLE CAN AND CANNOT SAFELY DRIVE

Orian Welling, PhD[1], Aaron Meyer, PhD[1], Sergey Vecherin, PhD[1], Michael Parker[1]

[1]US Army, ERDC-Cold Regions Research and Engineering Laboratory (CRREL), Hanover NH

ABSTRACT

Determining where a vehicle can and cannot safely drive is a fundamental problem that must be answered for all types of vehicle automation. This problem is more challenging in cold regions. Trafficability characteristics of snow and ice surfaces can vary greatly due to factors such as snow depth, strength, density, and friction characteristics. Current technologies do not detect the type of snow or ice surface and therefore do not adequately predict trafficability of these surfaces.

In this paper, we took a first step towards developing a machine vision classifier with an exploratory analysis and classification of cold regions surface images. Specifically, we aimed to discriminate between packed snow, virgin snow, and ice surfaces using a series of classical machine learning and deep learning methods. To train the classifiers, we captured photographs of surfaces in real world environments alongside hyperspectral scans, spectral reflectance measurements, and LIDAR. In this initial analysis, only the photography was assessed. The classifiers were cross-validated with a subset of the data collected for the project.

In addition to surface imagery, trafficability metrics were collected for each surface in the study. Vehicles from three different military classes (lightweight ATV, light, and medium/heavy) were tested as a modified Jeep Wrangler. Trafficability tests included draw bar and motion resistance, along with acceleration, deceleration, slalom, lane-change and circle dynamics tests where feasible. Each test surface was characterized alongside the vehicle and terrain sensing measurements and include measurements of density, temperature, and strength where applicable.

Results reported here show that winter surface conditions can be classified with 70%+ accuracy using onboard photography. Future work includes incorporating additional sensor data, vehicle, and snow data into the model.

Citation: Welling, O.Z., Meyer, A., Vecherin, S., Parker, M., "Determining Where a Vehicle Can and Cannot Safely Drive", In *Proceedings of the Ground Vehicle Systems Engineering and Technology Symposium* (GVSETS), NDIA, Novi, MI, Aug. 10-12, 2021.

1. INTRODUCTION

Determining where a vehicle can and cannot safely drive is a fundamental problem that must be answered for all types of vehicle automation and particularly for off-road vehicle operation where a wider range of conditions exist. This problem is made more challenging in cold regions by snow and ice. Trafficability characteristics of snow and ice surfaces can vary greatly due to rapidly changing factors such as snow depth, strength, density, and friction characteristics. Current vehicle proprioception technologies do not detect the type of snow or ice surface and therefore are not able to adequately predict trafficability of these surfaces. The US Army must improve these technologies to be better equipped for potential future conflicts in northern climates, and to enable Army Robotic and Autonomous Systems (RAS) strategy in these climates [1].

A long-term goal for cold regions vehicle automation is to develop a sensor system governed by artificial intelligence and/or machine learning (ML) algorithms to predict trafficability of a surface. A comprehensive solution should support navigation at scales from full route-planning to real-time vehicle reaction and control, and may include long range sensing, short range sensing and vehicle proprioception. Long range would use satellite and unmanned aerial systems or standoff sensing technologies to provide large scale terrain characterization, short range would use vehicle mounted sensors to provide an updated higher resolution vehicle perspective terrain characterization, and vehicle proprioception would allow the vehicle to determine physical or strength based parameters like the friction coefficient of the surface directly under the vehicle. All of this information can be fed into a neural network to update vehicle behaviors when this surface is encountered again. In this study, we focused on short range sensing and took a first step towards predicting the trafficability of a surface with an exploratory analysis and classification of cold regions surface images. Specifically, we aimed to discriminate between three winter surfaces (packed snow, virgin snow, and ice), sky, and vegetation using a series of machine learning and deep learning (DL) classification methods. The difference between the ML and DL methods is in identifying the characteristic features distinguishing between different terrain classes. For the deep learning neural networks, the raw, labeled images (with some elementary pre-processing) were used as a training set. Once trained, the neural network classified new images, without specifying what features distinguish one class from the other. For machine learning algorithms, a manual analysis of available images and data was performed first to identify the class-specific features, and then the identified features were used in the classification algorithm. A flow chart for the two approaches to classification are shown in Figure 1.

Both machine learning and deep learning techniques have been used successfully for terrain classification for autonomous vehicles in prior studies. However, to date, most tools developed for cold regions surface classification have been limited to simple determinations of snow cover. For example Nolte et al. [2] developed a deep learning classifier using vehicle based imagery to distinguish six types of terrain, of which “snow” was a single class (other classes include asphalt, dirt, grass, wet asphalt, and cobblestone). Similarly Taylor et al. [3] (also reported in Parker et al. [4]) used a machine learning classifier to make a binary distinction between snow-covered and bare ground and demonstrated how an autonomous vehicle could change operating parameters on the snow covered areas to improve performance. Rather than developing a classification, Elder et al. [5] used a field spectrum analyzer to collect reflectance signatures of different cold terrain surface types and compare

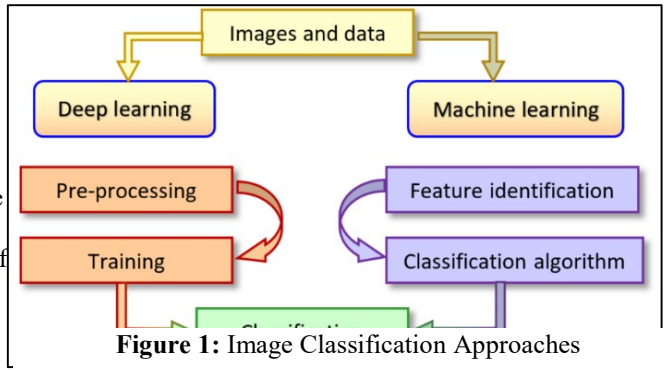


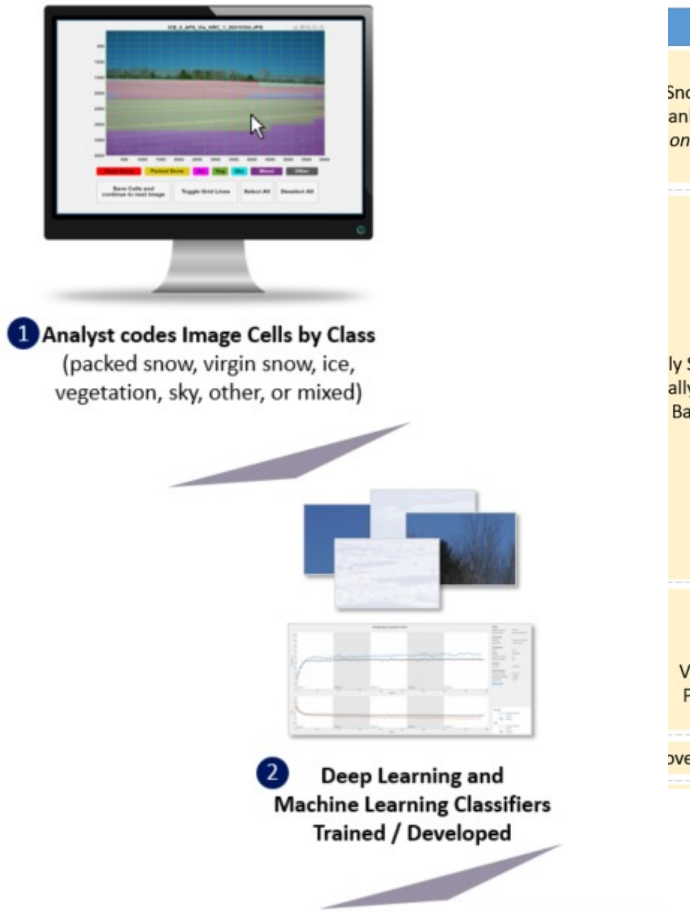
Figure 1: Image Classification Approaches

directly with snow strength [6] (an important physical property related to trafficability). The only existing example of a regional and more broadly applicable cold regions terrain classification for surfaces of interest for winter vehicle mobility, including packed snow, snow banks, virgin snow, and ice, is the work by Olivier & Shoop [7] who used satellite-based multi-spectral imagery rather than vehicle based imagery for their classification, so while it is helpful for route planning it offers only limited utility for on-the-ground autonomous driving. It is, however, feasible that satellite-based classifications could be used alongside real-time sensors to improve terrain classification accuracy for autonomy. Or that similar real-time imagery captured by UAVs could be fed directly into vehicle automation algorithms. Some work has already been done to leverage UAVs as a sensor platform for off-road autonomous vehicle navigation, such as the work by Vecherin et.al using UAVs to identify obstacles in a vehicles path [8].

In addition to surface classification for vehicle automation and route planning, there is substantial active research being conducted on road surface classification for the purpose of regional road surface monitoring. This research is similar in principal to the classification used in automation and the techniques could potentially be leveraged for automation applications. Road Weather Information Systems (RWIS) have become common in recent years [9]. These systems consist of weather stations, cameras, and frequently, pavement temperature sensors, and are used both to triage road networks for winter maintenance (snow removal, salting, and sanding) and to estimate road safety conditions (e.g. may be used to change variable speed limits or to inform road closures). Several researchers have developed computer vision tools using imagery from RWIS to estimate trafficability of roads. Yasuno et al. [10] developed a scalar snow rate hazard index related to the percent of a roadway covered in snow which they estimated based on RWIS and CCTV imagery using deep learning. Research at the University of Waterloo [11] developed a deep learning-based classifier to estimate a three-category winter road surface condition metric (RSC) from RWIS imagery in Ontario, Canada. The RSC classifications included bare pavement, partial snow cover, and full snow cover. Additional studies at the University of Waterloo showed the RSC metric could also be estimated from Smartphone [12] and dash cam [13] images and Shoop and Coutermarsh [14] used on-board vehicle sensors compared with onboard diagnostics derived data for use of vehicles as a winter road-weather sensors.

This study advances the field of computer vision for cold-regions surface classification and vehicle automation by extending vehicle based models to identify three types of winter surfaces with different trafficability metrics (ice, packed snow

Table 1: Machine learning and deep learning approaches used in prior snow classification efforts compared to this study



(PS), and virgin snow (VS)). The work builds on the prior research discussed above which successfully identified snow covered vs. non-snow covered terrain from ground and vehicle based sensors, and broader classifications with satellite imagery. A comparison of tools and classification capabilities between this and prior studies of imagery classification for cold terrain is shown in Table 1.

2. DATASET

Cold-regions surface data was collected in two locations for this study: Team O'Neil Rally School in Dalton, NH, and the Keweenaw Research Center in Calumet, MI. In both cases, image and spectral data were collected alongside vehicle tests and measurements of the surface density, temperature, and strength. The imagery and spectral data collected included RGB photographs, red / near-infrared photographs (using an 830nm band pass filter), and spectral reflectance measurements. At Team O'Neil, hyperspectral imagery and LIDAR were also collected.

	Visible	830nm
ICE	952	912
Mixed	1,191	1,026
Other	1,353	1,742
PS	7,325	6,242
Sky	2,782	1,953
Veg	1,354	1,068
VS	7,478	5,900
Total	***22,435***	***18,843***

For the preliminary analysis, only the RGB and 830nm filtered photography where processed. The 830nm filter was chosen (vs other filters) based on work from Wiscombe and Warren [15] that predicts the highest difference in reflectance of snow with different grain sizes in the near infrared bands (~800-2,500nm) – it was thought that differentiation of snow grain size would be likely to correlate with snow surface type. A total of 149 full images were captured (77 RGB and 72 830nm lens). Care was taken to collect images at a variety of conditions and angles: at each location images were taken horizontally, -30 degrees from horizontal, and directly toward the ground. Images were captured toward the sun and away from the sun, in cloudy and sunny conditions, and at different times through the day.

Image pre-processing included segmenting each original image into 20 x 20 grids and manually classifying each "cell" of the grid with one of seven classes (ice, packed snow, virgin snow, vegetation (Veg), sky, other, and mixed). To aid classification, the original images were taken at locations where the dominant surface type was uniform and co-located with vehicle testing. Images taken horizontally and at -30 degrees generally included multiple classes while the original images taken directly at the ground only contain a single class. The full count of image "cells" used in development of the classifiers is shown in Table 2 and the process for segmentation and classification process is shown in Figure 2

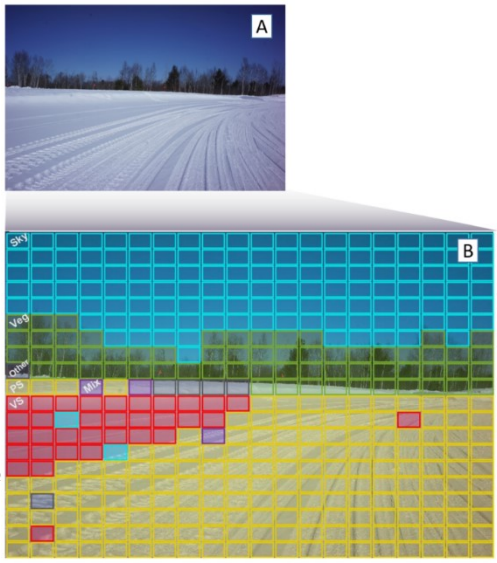


Figure 2: Images are segmented and each cell is classified prior to training the classifiers. The completed classifiers are also applied to new images at the cell level.

Table 2: Count of image "cells" available. In each test, the models were trained on 70% of the data by class, and 30% was held back for validation.

3. DEEP LEARNING APPROACH AND RESULTS

In our initial approach, we used transfer learning

to re-train three deep learning networks with the RGB and 830nm cell level imagery. The neural networks used included:

- AlexNet – A popular convolutional neural network (CNN) for computer vision.
- SqeezeNet – A computationally light-weight CNN designed similarly to AlexNet.
- InceptionNet V3 – A powerful, deep (and more computationally expensive) CNN commonly used in computer vision, and chosen for comparison to [11] and [2].

Each network was pre-trained using images from the ImageNet database. Only the last few classification layers were re-trained in each network. Thirty percent of each class of the snow surface images were excluded from the training and used for validation. The highest validation accuracy of 67.8% was achieved with AlexNet on classification of RGB imagery. Final validation accuracies of each network and imagery are shown in Table 3.

Table 3: Final validation accuracy of each deep learning network / imagery type.

	RGB Imagery	830nm Imagery
AlexNet	67.8%	54.4%
SqueezeNet	64.6%	54.7%
InceptionNet V3	61.7%	53.4%

Importantly, these data represent cell-level rather than full image level classification accuracy. For any given image, some cells will be classified incorrectly. However, regions of particular snow patterns may still be visibly obvious. For example, Figure 3 shows the result of applying the trained SqueezeNet to an RGB image. Some incorrect cells are clearly visible (e.g. two cells classified as "sky" in the virgin snow field), but general boundaries of VS and PS areas are easily observed.

A further understanding of the classification effectiveness can be gained by looking at the class-level accuracy. Table 4 shows the class-level accuracy for AlexNet using RGB imagery. Of the three winter surface categories, it can be seen that virgin snow classification is fairly accurate at 88%, while packed snow classification is lower at 66%, and ice classification is very low at 6%. The very low accuracy in classification of ice is troubling since it is a surface of significant importance in determining trafficability, and it may be related to the lower volume of ice images in the training set (only 4% and 5% of visible and 830nm cells respectively) compared to other classes. Overall, the best accuracy for classification of ice was 33% achieved with AlexNet applied to 830nm images, but the overall validation accuracy was lower. Future work, including adding additional ice images and/or sensor frequencies that better distinguish between ice and snow to the training data set, will be conducted to improve the classification accuracy of ice.

Figure 3: Example classification: SqueezeNet applied to RGB image. (A) Original RGB image. (B) Cell-level classified image with the trained neural network. While some cells are misclassified (e.g. two cells classified as "sky" in the virgin snow field), regions of different surface types are discernable.

A second observation from the class-level accuracies is the relatively poor performance in classification of "mixed" and "other" classes. These lower accuracies are not as surprising or problematic as the misclassification of ice: in the manual cell classification step, the "other" and "mixed" class were primarily used to separate anomalies from the main training categories. The "mixed" category, for example, was used if part of the cell was packed snow and part was virgin snow, and the "other" category was used for a wide variety of other objects such as people, road signs, building, and vehicles. In both cases, it is unsurprising that a cell may be miss-classified by the deep learning network due to the wide variability of these classes, and due to many of these cells containing partial areas of another class. To account for this, we also calculated the validation accuracy for all classes *excluding* "mixed" and "other". Again, AlexNet trained on RGB images achieved the highest adjusted

accuracy at 73.6% (SqueezeNet RGB 70.7%, InceptionNetV3 RGB 67.5%, AlexNet 830nm 61.3%, SqueezeNet 830nm 62.3%, and InceptionNetV3 830nm 61.4%).

Finally, it should be noted that some overfitting of the training data was observed while training each of the deep networks. Overfitting occurs when the network is able to distinguish features unique to the training images that, while representative of the specific training data, are not representative of the class. This overfitting can be observed as the classification accuracy of training data diverges from the validation accuracy through training as can be seen by the separation of blue and black lines in Figure 4. Overfitting can be mitigated by tailoring the training parameters and/or introducing randomness into the network

Actual \ Predicted	ICE	Mixed	Other	PS	Sky	Veg	VS	
ICE	53	3	18	359			520	6%
Mixed	2	322	11	430	20	137	270	27%
Other	3	131	251	151	348	273	197	19%
PS	6	534	20	4,842	15	2	1,907	66%
Sky		2	51	45	1,966	16	703	71%
Veg	18	27	16	57	23	1,194	20	88%
VS	2	76	6	631	166	6	6,592	88%

Percent of each category Correctly Classified

Total Accuracy: 67.8%

Accuracy excl. Mixed and Other: 73.6%

(e.g. randomly excluding different layers on different training runs). In this study, we were limited by computational time to iterate and optimize training parameters (this is left for future work). However, the maximum divergence between training accuracy and validation accuracy was less than 10% in every case, which was deemed acceptable for the initial analysis, and the accuracy on the validation set was leveled, and not decreasing.

Table 4: Class level accuracy of AlexNet trained on RGB imagery, counts show the percent of cells from data set aside for validation by their true and predicted (by AlexNet) classes.

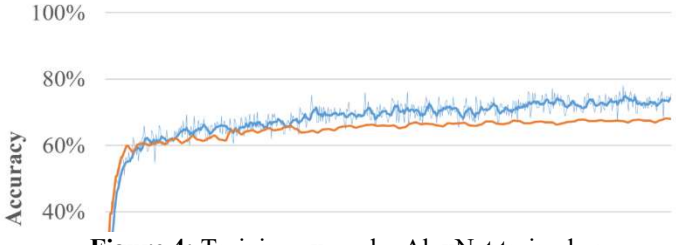


Figure 4: Training example: AlexNet trained on RGB imagery. Some overfitting is observable in the divergence between the blue (training accuracy) and black (validation accuracy) curves. However, the final observed difference was less than 10% which was deemed acceptable for this study

4. MACHINE LEARNING APPROACH AND RESULTS

The machine learning algorithm developed for this study is based on examination and comparison of statistical moments, calculated from the collection of pixels in each "cell" image. In contrast, other ML classifiers may focus on shape, or edge detection, especially if the classifier is designed to identify solid objects. The classification of snow and ice types is qualitatively different, there are no solid objects, which can be detected and classified. Therefore, this approach extends the texture filtration idea, which was deemed to be the most useful for snow type classification. In this section, two types of ML classifiers are developed, both based on this statistical approach.

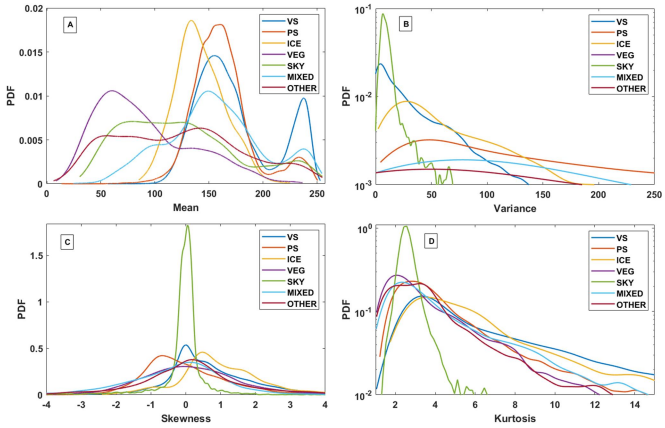


Figure 5: Probability density function (PDF) estimates of gray scale intensity Mean (A), Variance (B), Skewness (C), and Kurtosis (D), for all seven class types.

The initial step taken is conversion of the red-green-blue (RGB) cell images to gray scale images representing the light intensity. The conversion to intensity is given by the weighted average in Equation 1:

$$Y = 0.299R + 0.587G + 0.144B \quad (1)$$

where *R*, *G*, and *B* are integer numbers ranging from [0 255] that describe the red, green, and blue color in each pixel, and *Y* is the corresponding intensity. This transformation corresponds to the YIQ color space used by the National Television System Committee (NTSC), and approximates the black and white color imagery used by mid-20th century American televisions [16]. After conversion to gray scale via Equation 1, each cell image is composed of a distribution of intensity values.

Next, the first four statistical moments (mean, variance, skewness, and kurtosis) of these distributions are calculated and used as metrics for image classification. For brevity, these metrics will be abbreviated to *M*, *V*, *S*, and *K*. Each cell image in the data set is now described by these four numbers. The next step is to aggregate the statistical moments by class (VS, PS, Ice, Veg, Sky, Mixed, Other) and observe the distribution of these metric values within the entire data set. Figure 5 (A-D) show probability density function (PDF) estimates of class-specific *M*, *V*, *S*, and *K* distributions, respectively.

The PDF estimates were obtained by using a kernel density estimate (KDE) with Gaussian kernel and optimal smoothing bandwidth parameter. The bandwidth parameter is an ingredient in the PDF calculation which governs the width of the kernel function, in this case, Gaussian function. Because the classifier depends on PDF estimates, its performance is sensitive to the choice of bandwidth parameter. The bandwidth chosen for this application is given by Equation 2:

$$k = 1.06\,\sigma\, n^{-1/5}, \qquad (2)$$

where σ is the standard deviation of the respective *M*, *V*, *S*, and *K* distributions, and *n* is the number of samples. The compact Equation 2 is known as "Scott's rule of thumb" [17].

For a given image with a set of values (*M*, *V*, *S*, *K*), the values of class-specific PDFs seen in Figure 5 characterize likelihood that the image belongs to the specific class. For example, consider only Figure 5(A), which shows the class-specific PDF estimates of the average intensity values (*M*). Suppose, an image is randomly chosen from the data set and found to have an *M* value of 150. Referencing the PDFs in Figure 5A, one sees that for *M* = 150, the PDF corresponding to packed snow (PS) has the highest value, followed by VS and Ice, then Mixed, and so on. Therefore, based on the *M* metric, the randomly sampled image most likely belongs to the PS class, followed by VS and Ice, then Mixed, etc. This logic is equally applied to the other three statistical metrics (*V, S, K*) and underpins the first ML algorithm developed for this study. Let us denote the PDF value of an image *i*, as $P(x_i|Q,C)$. Here, Q defines one of the four statistical metrics (*M*=1, *V*=2, *S*=3, *K*=4), x_i is the *i*th image PDF argument for that metric, and the condition *C* denotes one of the seven classes (VS=1, PS=2, Ice=3, Veg=4, Sky=5, Mixed = 6, Other = 7). As an example, the quantity $P(x_i|2,5)$ gives the probability density estimate for the variance of the image *i*, given the distribution of variances contained in the Sky class. This quantity indicates the likelihood that the image *i* belongs to the Sky class. In total, there are twenty-eight combinations of Q and *C*, which correspond to the twenty-eight PDFs seen in Figure 5. The first classifier is then defined by Equation 3:

$$I(i) = \underset{\{C\}}{\operatorname{argmax}} \prod_{Q=1}^{4} P(x_i|Q,C), \qquad (3)$$

where *I(i)* is an index indicating the classification of the image *i*. The four-fold product is tantamount to the "AND" operation in probability theory, i.e. for the image *i* to belong to class C, all metrics should be sufficiently high, *M* AND *V* AND *S* AND *K*.

The classifier described by Equation 3 on the gray scale jpg imagery achieved a training set accuracy of approximately 51% overall, with individual class accuracy of 47.2% for VS, 53.3%

for PS, 60.1% for Ice, 65% for Veg, 79.4% for Sky, 23.4% for Mixed, and 11.5% for Other. 'Training set' in this context means all images were used to calculate the PDFs, and tested by the classifier. Training-validation simulations were also conducted, where a fraction of the images are used to train (calculate the PDF) and the remaining images used to validate. The training-validation simulations using this type of classifier were less accurate (< 51%) than the training set. Although the performance of the first classifier is unimpressive, it provided "proof of concept" in the PDF approach, and motivated the development of the second type of machine learning classifier used in this study.

The second type of classifier is based on multivariate PDFs. The PDFs seen in Figure 5 are one-dimensional, but it is possible to calculate N-dimensional, or multivariate PDFs, using N-dimensional kernel functions. In this approach, the four statistical moments (*M*, *V*, *S*, *K*) now represent each dimension, and individual images can be conceptualized as inhabiting that 4-dimensional space. In addition, the information about the color (*R*, *G*, *B*) of each cell image can be incorporated into this multidimensional framework. Figure 6 shows a 3-dimensional scatter plot which maps where each of the first five classes (VS, PS, Ice, Veg, and Sky) exist in the (*R*, *G*, *B*) space. The coordinates of individual scatter points are calculated as the mean of the R, G, and B pixel values comprising each cell image.

The class specific cell images produce a nontrivial spatial distribution in the (*R*, *G*, *B*) space. Typically for automatic classification, the effectiveness of this type of classifier depends on good spatial separation of the classes. For instance, a good separation between the points representing Sky images and Vegetation images can be observed in the *Blue* dimension in Figure 6.

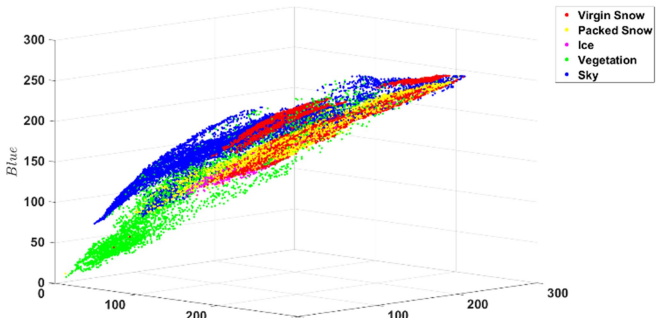


Figure 6: Three dimensional scatter plot of mean R, G, B color for images in the first five class types.

Let us denote a multivariate PDF value of an validation image *i* as $P_{val}(\mathbf{x}_i|\mathbf{Q},C)$. Here a subscript *val* is introduced to indicate that the image *i* belongs to a validation set of images, and is therefore classified using PDF estimates calculated from a different sub set of images. The index $\mathbf{Q}$ now indicates a list of statistical metrics, whereas Q previously only indicated one metric. There are up to seven metrics to consider: (R = 1, G = 2, B = 3, M = 4, V = 5, S = 6, K = 7). As an example, the quantity $P_{val}(\mathbf{x}_i|[2\ 5\ 6],3)$ gives the multivariate PDF estimate of the image *i*, in the 3-D space defined by the (G, V, S) metrics, trained by images belonging to the Ice class. The second classifier is then defined by Equation 4:

$$I(i,\mathbf{Q}) = \underset{\{C\}}{\operatorname{argmax}}\, P_{val}(\mathbf{x}_i|\mathbf{Q},C), \qquad (4)$$

where $I(i,\mathbf{Q})$ is an index indicating the classification of the image *i* with the set of metric values $\mathbf{x}_i$, in the multi-dimensional space defined by the list of metrics, $\mathbf{Q}$.

The accuracy results for the second type of classifier is summarized in Table 5. Here, the analysis was conducted for both the visible, and 830 nm images, indicated by the first column of Table 5. Three combinations of metrics were tested, RGB, MVSK, and RGBMVSK, indicated by the second column.

Table 5: Overall and class level accuracy (percent) for the multivariate PDF classifier type.

Spectrum	Metrics	VS	PS	Ice	Veg	Sky	Mixed	Other	Total Accuracy	Accuracy excl. Mixed & Other
Visible	RGB	31	79	54	54	68	4	1	**50.4**	**56.6**
Visible	MVSK	62	61	73	77	81	30	19	**60.7**	**65.4**
Visible	RGBMVSK	60	77	73	74	77	16	8	**63.7**	**70.4**
830nm	RGB	4	89	43	16	50	0	3	**35.5**	**45.0**
830nm	MVSK	68	50	66	55	75	30	25	**56.0**	**61.2**
830nm	RGBMVSK	64	75	56	44	74	7	12	**58.6**	**67.5**

The accuracies shown in Table 5 are percentages, averaged over one hundred 70-30 training-validation simulations with random selection of images for the training and validation sets. During one 70-30 training-validation simulation, the collection of images for each class

type are proportionally split into two sets: a training set comprised of 70% of all the images, and a validation set comprised of the remaining 30%. Determination of which images belong to which set is randomized for each simulation. The randomization creates variability in the resulting accuracy of classifications for a given 70-30 training-validation simulation, hence the need to average over many (100) such simulations. The accuracy estimates seen in Table 5 are color coded. Classifications with >70% accuracy are shaded dark green, 70-60% are light green, 60-40% are orange, and <30% are red.

Several insightful conclusions can be drawn from Table 5. First, the classifier delivers generally better results for visible spectrum jpg images than 830 nm imagery. This is in agreement with the deep learning networks. There are a few exceptions to this rule, notably the classification accuracy of packed snow with RGB is greater for 830 nm than visible, by an appreciable margin (~ +10%).

Second feature to note is the universally poor classification accuracy of the Mixed and Other classes, which was discussed previously in the deep learning section. Next, it is worth noting that for all but one class, the classifier using MVSK information performs better than RGB. This seems to confirm the underlying hypothesis behind the use of the statistical ML approach. MVSK metrics are representative of the "texture" or "roughness" of the images, and are more useful for classification than other features, such as color. The exception to this rule is packed snow, which is classified with a higher accuracy using information about color in both the visible and 830 nm spectrums.

Next, note that "adding in dimensions" to the classifier, e.g. comparison of RGB to RGBMVSK does improve total classification accuracy, indicated by the last two columns in Table 5, but is generally not true for individual classes. For example, the classification accuracy for virgin snow (VS) in the 830 nm spectrum is very poor (3.5%) for RGB, but significantly improved for MVSK (68.1%). However, the accuracy when all seven metrics are used, RGBMVSK, provides worse classification accuracy (63.8%). Qualitatively, it seems the reintroduction of information about the color polluted the otherwise better classification from MVSK alone.

Last, note that the reduced accuracy of the RGBMVSK classifier compared to only MVSK is true for all classes except PS, and yet the total accuracy for RGBMVSK is better than MVSK. This is due to the PS class having a disproportionally larger share of the image data set. For the visible images, the PS class accounts for ~33.2% of all images. Therefore the total accuracy, which is calculated as the ratio of all images classified correctly to all images tested by the classifier (regardless of class), is affected by PS images more than other classes.

The PDFs used in the second type of classifier were calculated using the multi-dimensional generalization of Equation 3 [18] for the kernel bandwidths:

$$k = \sigma \left(\frac{4}{n(d+2)} \right)^{\frac{1}{d+4}}, \quad (4)$$

where σ is the standard deviation of the respective one dimensional distributions, n is the number of samples, and d is the number of dimensions of the multivariate PDF. As mentioned previously, the performance of the PDF classifiers is sensitive to the selection of bandwidth parameters.

Investigation of varying the bandwidth parameters, k, revealed that further classifier accuracy could be gained by choosing larger k than what is predicted by Equation 4. Larger k effectively makes the kernel functions wider, and the resulting multivariate PDFs smoother by appearance. Table 6 shows results for the second type of classifier with "best guess" choices made for the bandwidth parameters. The improvements to accuracy, compared to Table 5, are significant and highly encouraging. This result suggests an

Table 6: Overall and class level accuracy (percent) of the multivariate PDF classifier type with manual bandwidth parameter adjustment.

Spectrum	Metrics	VS	PS	Ice	Veg	Sky	Mixed	Other	Total Accuracy	Accuracy excl. Mixed & Other
Visible	RGB	73	81	74	86	70	34	29	71.4	76.5

optimization of bandwidth parameters is possible, and warrants further investigation.

5. DISCUSSION AND RESULTS

Both deep learning and traditional machine learning approaches showed promising results in this initial analysis. Overall, the machine learning approach with manual adjustment of the bandwidth parameter showed the best results at 71.4% classification accuracy for RGB imagery. AlexNet was close behind with 67.8% accuracy for RGB imagery. Notably, the machine learning approach was much more accurate at classifying ice than even the best case deep learning approach (74% vs. 33% for AlexNet classification of 830nm imagery). This may be a related to the machine learning focus on textural image properties vs. the deep learning networks which were each developed and pre-trained for object recognition. Additionally, although not implemented in this paper, the machine learning approach provides an opportunity to directly optimize and/or weight specific categories of interest while the deep learning approach in general does not.

There remains significant opportunity to further refine both deep learning and machine learning approaches to increase classification accuracy. On the algorithm side, there are several deep learning networks that have been developed specifically to decode textural rather than shape data (such as networks developed for interpretation of x-ray images). The authors believe that these networks may perform better than the more common neural networks tested here. On the machine learning side, it is possible to further expand the developed approach to account for more than 7 variables which may increase accuracy. For both approaches, removing “other” and “mixed” classes from the training data set entirely may reduce false negatives for the packed snow, virgin snow, and ice classes that are most important for mobility.

Possibly more important that algorithmic improvements are refinements to the data collected and analyzed. Oliver and Shoop [7], for example, achieved 95.8% classification accuracy, with similar classes, when analyzing short-wave infrared (SWIR) satellite imagery instead of only RGB imagery. In addition to the RGB and 830nm imagery discussed in this analysis, the authors collected hyperspectral imagery, LIDAR, and spectral reflectance that may each have advantages over traditional imagery. As a next step the authors will analyze the additional sensor data to determine if a higher classification accuracy can be achieved from other sensors. The authors will also mix-and-match data through sensor fusion to determine if there is a set of multiple signatures that significantly outperforms the individual sensors. Once an optimal set of sensors is identified, the authors will work to further segment the class structure to capture additional snow types that impact mobility (in particular segmenting the virgin snow class into multiple sub-classes). The end goal will be to determine a combination of a small number of sensors that could be developed into a terrain sensing package suitable to integrate with an autonomous vehicle control system to improve off road performance in cold regions.

1. REFERENCES

[1] US Army: TRADOC Maneuver, Aviation, and Soldier Division Army Capabilities, "The U.S. Army Robotic and Autonomous Systems Strategy," U.S. Army Training and Doctrine Command, Fort Eustis, VA, 2017.

[2] M. Nolte, N. Kister and M. Maurer, "Assessment of Deep Convolutional Neural Networks for Road Surface Classification," in *21st International Conference on Intelligent Transportation Systems (ITSC)*, Maui, HI, 2018.

[3] T. S. Hodgdon, A. J. Fuentes, J. L. Olivier, B. G. Quinn and S. A. Shoop, "Automated Terrain Classification for Vehicle Mobility in Off-Road Conditions," ERDC/CRREL TR-21-4, Hanover, NH: U.S. Army Engineer Research and Development Center, 2021.

[4] M. Parker, B. Quinn, J. Bates, T. Hodgdon, M. Bodie, S. Shoop and A. Stott, "Exploring cold regions autonomous operations," *Journal of Terramechanics,*

2021.

[5] B. Elder, S. A. Shoop and M. Michaels, "Spectral Reflectance Signatures of Compacted Snow Surfaces," in *Eastern Snow Conference Procedings*, Lake Morey, VT, 2019.

[6] S. A. Shoop, W. Wieder, B. C. Elder, S. Beal and E. Deeb, "Assessment of Field Methods for Measuring Mechanical Properties of Snow," ERDC/CRREL TR-19-17, Hanover, HJ: U.S. Army Engineer Research and Development Center, 2019.

[7] J. Olivier and S. Shoop, "Imagery Classification for Autonomous Ground Vehicle Mobility in Cold Weather Environments [PLACEHOLDER]," *ISTVS,* 2021.

[8] S. N. Veherin, J. M. Shaker and M. W. Parker, "Obstacle Detection and Quantification for Vehicle Mobility in Winter Conditions," ERDC/CRREL TR-20-7, 2020.

[9] S. E. Boselly III, G. S. Doore, J. E. Thornes, C. Ulberg and D. D. Ernst, "Road Weather Information Systems Volume 1: Research Report," Strategic Highway Research Program, Washington, DC, 1993.

[10] T. Yasuno, J. Fujii, H. Sugawara and M. Amakata, *Road Surface Translation Under Snow-covered and Semantic Segmentation for Snow Hazard Index,* vol. 2101.05616, 2021.

[11] J. Carrillo, M. Crowley, G. Pan and L. Fu, "Design of Efficient Deep Learning models for Determining Road Surface Condition from Roadside Camera Images and Weather Data," in *Artificial Intelligence and Machine Learning for Smart Mobility Session of the 2019 TAC-ITS Canada Joint Conference*, Halifax, NS, 2019.

[12] G. Pan, L. Fu, R. Yu and M. Muresan, "Winter Road Surface Condition Recognition Using a Pre-trained Deep Convolutional Neural Network," in *Transportation Research Board 97th Annual Meeting*, Washington, DC, 2018.

[13] M. Linton and L. Fu, "Connected Vehicle Solution for Winter Road Surface Condition Monitoring," *Journal of the Transportation Research Board,* vol. 2551, pp. 62-72, 2016.

[14] S. Shoop and B. Coutermarsh, "Vehicles as Weather and Road Condition Sensors," in *International Symposium on Cold Regions Development*, Tampere, Finland, 2007.

[15] W. J. Wiscombe and S. G. Warren, "A Model for the Spectral Albedo of Snow. I: Pure Snow," *Journal of the Astropheric Sciences,* pp. 2712-2733, 1980.

[16] International Telecommunication Union, "Characteristics of composite video signals for conventional analogue television systems," 2004.

[17] D. Scott, "Theory, practice and visualization," NY: John Wiley & Sons, 1992.

[18] A. W. Bowman and A. Azzalini, "Applied Smoothing Techniques for Data Analysis: The Kernel Approach with S-Plus Illustrations.," Oxford University Press, Oxford, England, 1997.