

Classification of 3D Point Cloud Data from Mobile Mapping System for Detecting Road Surfaces and Potholes using Convolution Neural Networks

Chaithavee, S. and Chayakul, T.*

Department of Survey Engineering, Faculty of Engineering, Chulalongkorn University, Phayathai Road, Wangmai, Pathumwan, Bangkok 10330, E-mail: Sorasak.rlee@gmail.com, Thongthit.c@chula.ac.th

*Corresponding Author

DOI: <https://doi.org/10.52939/ijg.v18i6.2453>

Abstract

Normally, road damages can be automatically detected using image and video data from ground survey vehicle system combined with the detection algorithms. However, there are limitations of scales and map coordinates when using the image and video data to detect potholes. It has been challenging to detect and determine the sizes and locations of potholes. This research utilized a mobile mapping system, MMS, to collect data of roads and environment and classify potholes, roads and other objects. A convolution neural network (CNN) was used to directly identify 3D point clouds using the XYZ method in comparison with the proposed XYZ-RGB method. The XYZ classification demonstrated an overall accuracy of 96.77%, with the intersection over union (IoU) of potholes, roads, and other objects of 59.50%, 94.22%, and 94.06%, respectively. The proposed XYZ-RGB classification indicated an overall accuracy of 97.50%, with the IoU of potholes, roads, and other objects of 66.66%, 95.43%, and 95.42%, respectively. Both datasets were statistically compared at the 95% confidence level, and the results revealed that both classifications produced significantly different results.

Keywords: Mobile Mapping System, Point Cloud, Classification, Convolution Neural Networks

1. Introduction

Road damage can negatively impact commuters and increase the risks of road accidents. Therefore, data on the physical conditions of road surfaces and damages are crucial and must be obtained for further applications such as planning, maintenance, and budget allocation. A systematic and continuous survey of road surface conditions is necessary to increase the road's level of service and service life. Therefore, it is essential to survey for road damage. One of the easiest methods is a ground survey, which requires experts to manually go on foot to assess by visual inspection. However, this conventional approach is moderately costly, requires several workers, consumes excessive time, and provides less reliable data. It also exposes surveyors to potential danger while working on the road [1]. Consequently, modern technology and equipment are used with automation to mitigate the risk associated with the conventional method of detecting road damage. Road surveys produce high efficiency and accuracy with technology, enabling effective surface analysis. Nonetheless, since road damage does not share a

definite characteristic, using technology to detect road damage might not be as effective as expected. Consequently, technology has rarely been used for automated road damage detection [2].

Road damages, especially potholes, can be detected automatically by analyzing photographic or video data derived from vehicle cameras in conjunction with the development of detection algorithms [3], [4], [5]. Machine learning (ML) is a popular choice of an instrument employed to detect potholes from photographic data. It has been implemented through several techniques such as Support Vector Machine (SVM) [6], [7], Random Forest (RF) [8], and neural networks [9]. In fact, neural networks, especially the Convolutional Neural Network (CNN), are the current trend in modern photographic pothole detection [10], [11], [12]. However, when using photographic or video data to detect potholes, there remain limitations related to scales and map coordinates. More specifically, it has been challenging to determine the precise sizes and locations of detected potholes.

Although unmanned aerial vehicle (UAV) imaging survey can detect road damages [13], [14] from scaled images in the map coordinate system and measure pothole sizes from aerial photographic map data, the method cannot yet detect or classify potholes in three dimensions (3D). As a result, the mobile mapping system (MMS) [15], [16] is more frequently utilized to detect potholes and other road objects in 3D since resulting point clouds can provide thorough details of road objects with positional accuracy.

MMS employs several sensors, including the Global Navigation Satellite System (GNSS), the Inertial Measurement Unit (IMU) sensor, 360-degree panoramic cameras, the Light Detection and Ranging (LiDAR) laser scanner, and the Distance Measurement Instrument (DMI) [17]. Mounted on the vehicle, MMS utilizes these sensors altogether to explore the area while moving. Collected data processed by MMS can be stored as a point cloud, a spatial resolution of point data where every point refers to a coordinate in the map system. Hence, several studies use point cloud data to classify objects in the road environment, such as road surfaces, electric poles, trees, cars, buildings, and traffic signs [18], [19], [20]. Since modern MMS is equipped with 360-degree panoramic cameras, point clouds are recorded in the color values of red, green, and blue (RGB), allowing 3D renders to show color values reflecting original photographic data and adding dimensions to point cloud classification [21]. Nevertheless, no studies have yet compared the accuracy of point cloud classification with and without RGB values.

Algorithms have been developed to detect and classify objects in the road environment [18] [20] and potholes [15], [16], [22] from point cloud data, and ML can learn to classify objects from point clouds and photographic data. In fact, several ML techniques have been utilized to classify trees, electric poles, buildings, traffic signs, and road surfaces. The primary ones include RF for road edges and traffic signs on road surfaces [23], SVM for objects in urban areas [24], and CNN for objects from point clouds [25].

Classifying point clouds with CNN is more complex than classifying with two-dimensional (2D) photographic data due to the disordered arrangement of points. CNN point cloud classification can convert 3D data into pixels [26] or voxels for extensive 3D-shape utilization [27]. Some CNN architecture, such as PointNet, has evolved to the extent that it can classify point cloud data directly [28].

PointNet is a highly functional architecture as it can directly classify point clouds and simplify the

process of converting data to voxels and pixels. Furthermore, the architecture has recently been upgraded to PointNet++ [29], with a distinctive feature allowing the architecture to detect objects' local regions more effectively. However, since point cloud data only contain XYZ point values, using CNN to study and classify them was suspected to be problematic because road surfaces and potholes are not as significant in levels, and most pothole areas have shallow depths. Consequently, the said practice might produce excessive classification errors. On the other hand, combining point cloud data with image data obtained from 360-degree panoramic cameras provides access to both the XYZ and RGB values, and RGB color values are helpful when distinguishing objects. Hence, this study explored the classification of point cloud data obtained from MMS using CNN via the PointNet++ architecture to detect roads and potholes. Furthermore, it compared the efficiency of the two-point cloud classification methods, which utilized XYZ-only and XYZ-RGB data, in their training, validation, and classification processes. The purpose was to assess the significance and usefulness of these RGB values in enhancing data classification.

2. Materials and Methods

2.1 Study Area

The spatial scope is Yothathikan Road, Keha Karnkaset Village, Village No. 10, Nong Prao Ngai Subdistrict, Sai Noi District, Nonthaburi Province, Thailand (N 13°52'48.37", E 100°18'37.51"), which is as shown in Figure 1. The road deserves exploring because it leads to several industrial plants. Moreover, with trucks heavily utilized for transportation, the road surface is prone to damage and contains several potholes. Consequently, this study employed MMS to survey approximately four kilometers of the road, and collected data were processed by CNN to classify roads and potholes for further damage assessment.

2.2 Mobile Mapping System

MMS has a rapid pace of development and is an advanced surveying and mapping technique that can efficiently and quickly collect spatial data. [30] Typically, MMS vehicles are mounted with sensors such as LiDAR, GNSS, IMU, DMI, and advanced digital cameras (Figure 2). They are also equipped with a central computer system for data storage and management. [31] In addition, since GNSS, IMU, and DMI are positioning and orientation systems (POS), mobile mapping is conducted by scanning for wavelength and detecting light intensity when the laser is reflected from object surfaces.

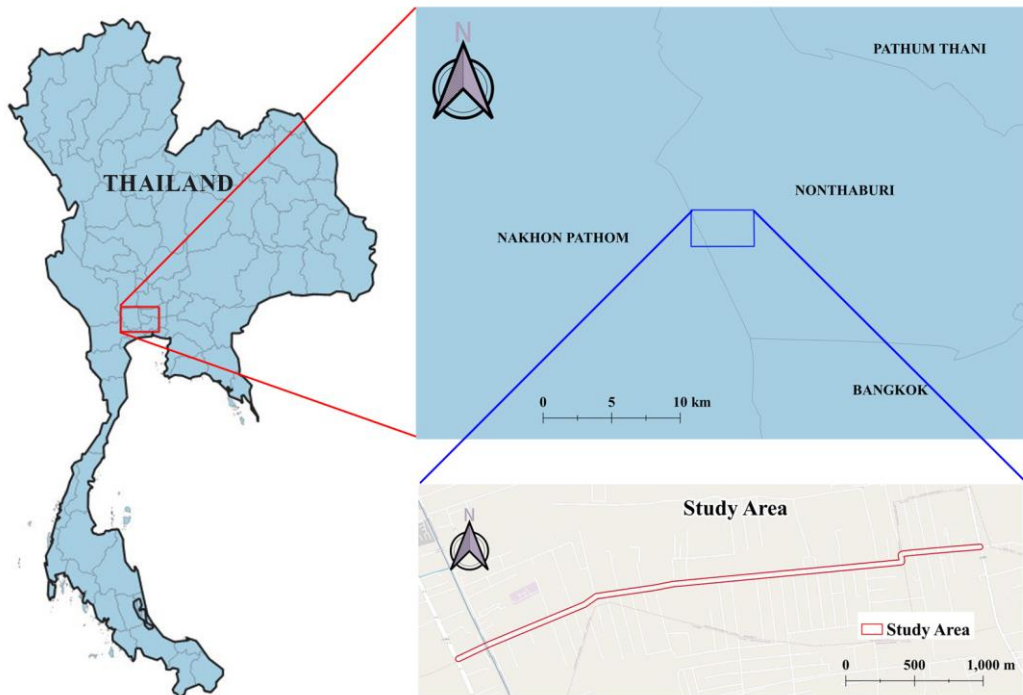


Figure 1: The location of Yothathikan Road, Keha Karnkaset Village, Village No. 10, Nong Prao Ngai Subdistrict, Sai Noi District, Nonthaburi Province, Thailand



Figure 2: MMS-IP-S3 mounted on a vehicle

MMS can scan objects with precision, and acquired data can be used to create a 3D model of a city [32]. Also, distance can be accurately measured by the speed of light, which generates a wave signal back and forth. Therefore, the positional accuracy of the 3D model depends on the angle at which the wave signal is scanned, measuring distance, and the

position and orientation of the device [33]. Figure 3 describes the mechanism in which map coordinates are referenced based on the scan angle α and the scan distance d of point P , which is determined by the coordinate system. In addition, position values in the scan coordinate system can be converted to coordinates in the map coordinate system.

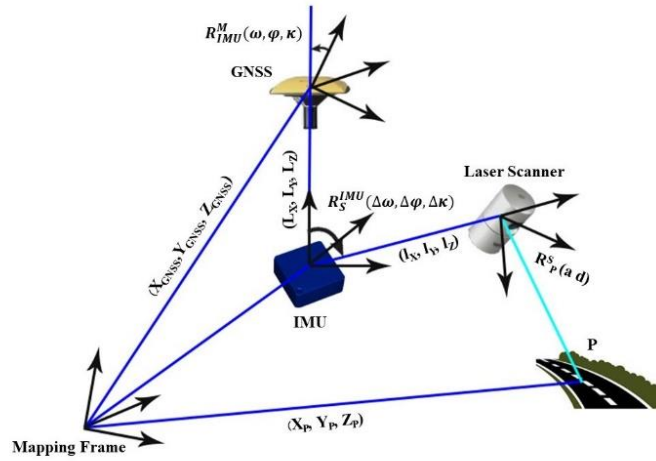


Figure 3: MMS's mechanism for referencing map coordinates

$$\begin{bmatrix} X_P \\ Y_P \\ Z_P \end{bmatrix} = R_{IMU}^M(\omega, \varphi, \kappa) \cdot \left(\begin{bmatrix} l_X \\ l_Y \\ l_Z \end{bmatrix} + \begin{bmatrix} L_X \\ L_Y \\ L_Z \end{bmatrix} + r_P^S(a, d) \cdot R_S^{IMU}(\Delta\omega, \Delta\varphi, \Delta\kappa) \right) + \begin{bmatrix} X_{GNSS} \\ Y_{GNSS} \\ Z_{GNSS} \end{bmatrix}$$

Equation 1

As shown in Equation 1, $[X_P, Y_P, Z_P]^T$ indicates the position value of the target object P in the map coordinate system, and $R_{IMU}^M(\omega, \varphi, \kappa)$, including the IMU-derived local mapping frame's sensor roll, pitch, and yaw, serves as the IMU-mapping rotation matrix. Furthermore, the navigation and the IMU origin obtained from a system calibration or a measurement provided the lever-arm offsets of $[l_X, l_Y, l_Z]$, which were equivalent to the origins of the laser scanner and GNSS. Additionally, laser scanner coordinates were determined by $r_P^S(a, d)$, the relative position vector of Point P. The laser scanner's angle and range were represented by a and d , whereas $R_S^{IMU}(\Delta\omega, \Delta\varphi, \Delta\kappa)$ served as the scanner-IMU rotation matrix, with $(\Delta\omega, \Delta\varphi, \Delta\kappa)$ indicating the scanner frame's boresight angles obtained through system calibrations and adjusted to match the IMU body frame. $[X_{GNSS}, Y_{GNSS}, Z_{GNSS}]^T$ indicates the location value of the GNSS receiver in the same map system

2.3 Using the Convolutional Neural Network in Point Cloud Classification

Since most CNN architectures take voxel and pixel data as input, point clouds are not native to them. Consequently, most researchers initially convert point cloud data to voxels or pixels [26], [27] before inputting it to CNN for processing. Unfortunately, data conversion often creates unnecessary data and degrades the quality of point cloud data. Therefore, if point cloud data could be classified directly, there would be no need for data conversion, removing the point cloud's unique characteristics and detailed

parameters. For instance, if RGB values are to be used to classify data in conjunction with XYZ coordinates, users can only choose to keep one of the parameters when converting the data, preventing the effective use of point cloud data. Fortunately, PointNet is a CNN architecture that can directly classify point cloud data [28]. Developed in 2017, this architecture has become a foundation for several new architectures. Furthermore, it can directly import point cloud data using an equation of $P=R^{N \times D}$ where N represents the number of point clouds and D is the dimension of the data. Normally, $D=3$ as the dimension (dim) refers to the XYZ values of each point.

However, with PointNet, RGB data can be used in conjunction with this dimension D. In addition, PointNet has been developed extensively and named PointNet++ based on CNN architectural principles. PointNet++ contains three main data layers: sampling, grouping, and PointNet. Initially, the sampling layer selects a set of points as centroids of the local regions. Subsequently, the grouping layer searches for neighboring points of these centroids. This layer receives a data input in the form of a point set of size $N \times (d + C)$ that is mainly composed of N points with C-dim point features, d-dim coordinates, and an array of coordinates for centroids of size $N' \times d$. Moreover, it generated point set outputs that were of the size $N' \times K \times (d + C)$, given that each of these groups must be consistent with a local region and K represents the quantity of points neighboring centroid points. Eventually, the PointNet layer employs mini-Pointnet to encode local region patterns into feature vectors.

More specifically, this layer receives a data input of N' local regions with points indicating data size $N' \times K \times (d + C)$. Furthermore, its output's local regions are abstracted by their centroid and local features encoding the centroids' neighboring areas. The data size of the output refers to $N' \times (d + C)$ matrix of N' subsampled points with d -dim coordinates and new C -dim feature vectors, which described the local context. The way to solve this problem is to sample all points as centroids at an inevitably higher computational expense at the abstraction level. An alternative is to propagate subsampled points' features to their original points. Hence, this study adopted the hierarchical propagation strategy using interpolation based on distances and across level skip links (Figure 4). Hence, point features were propagated from $N_l \times (d + C)$ to N_{l-1} where N_{l-1} and N_l (with $N_l \leq N_{l-1}$) represent the point set size of the input and output of set abstraction level l . In terms of interpolation, the inverse distance weighted average method was implemented in conjunction with the k -nearest neighbors algorithm (k -NN). Furthermore, skip-linked point features from the defined abstraction level are concatenated with the interpolated features on N_{l-1} points. Afterward, a "unit pointnet," comparable to a CNN's one-by-one convolution, was applied to the concatenated features. The feature vector for each point was updated by applying some shared fully linked and ReLU layers. The original point set was subsequently updated through the propagation of these features in a repeated process [29].

Open3D is an open-source library that supports the development of software related to 3D data. It is responsible for presenting structures and algorithms in C++ and Python; managing data for PointNet++ to load, write, and display point cloud data; and pre-processing, down-sampling, and interpolating data.

2.4 Data Collection

In this study, Topcon IP-S3 was utilized to handle MMS tasks by collecting and storing data of the study area, which is approximately a four-kilometer distance from the local control point, GNSS 610373. Obtained data were processed by the Post Process Kinematic (PPK), while six ground control points (GCPs) and six checkpoints (CPs) were measured to improve data accuracy. According to the National Standard for Spatial Data Accuracy (NSSDA), the model produced 0.089 meters of 3D accuracy at the 95% confidence level (NSSDA Class 2), suitable for mapping missions involving automatic and semi-automatic object classification [34]. The point cloud data derived from MMS contains approximately 282 million points, roughly at the density of 5,000 points

per square meter, with enough spatial resolution to classify potholes and roads. Furthermore, since the data were embedded with color values from digital cameras, each point contains both the XYZ and RGB values.

2.5 Point Cloud Classification through a Convolutional Neural Network

In this study, since CNN via the PointNet++ architecture was used to classify roads and potholes, point cloud data could be processed directly in two methods, including XYZ only and XYZ-RGB. Figure 4 illustrates the addition of RGB values when classifying roads and potholes from point cloud data. To prepare datasets for training and validation, in areas where data were unclear, potholes, roads, and other objects were manually labeled by visual observations via software, and the effort was supported by data from additional field surveys. These datasets were also utilized as references for checking the test datasets of both methods for classification accuracy (Figure 4). The trained model demonstrated that it could accurately classify objects from unknown data, the primary target for this classification. In other words, if the model could only classify previously known data, the network did not learn different information. Hence, training and validation are essential to learning, allowing models to predict unknown data more accurately. Furthermore, when classifying data, it is crucial to define conditions and parameters to suit the data appropriately. The training dataset contains three-dimensional (3D) point-data coordinates, and it was used to train CNN to enhance its classification and recognition capabilities involving data attributes. Simply put, the training dataset taught CNN to create classification models corresponding to its preset data attributes. Furthermore, the validation dataset also served as another set of data attributes that contributed to the mentioned post-training models. They were also used to test the classification, and their results were further incorporated to revise and enhance the models. Eventually, the testing dataset was employed as actual classification test data for the models since these models were never exposed to the test data before.

Hence, if the models worked effectively with the test data, they would also be adequately accurate to classify other data. Typical ratios of training, validation, and testing data are either 80:10:10, 70:15:15, or 60:20:20, depending on data characteristics [35]. In this study, the distribution of potholes was uneven distance-wise. Therefore, the 60:15:25 ratio was chosen because there were few potholes in some data ranges in the testing dataset.

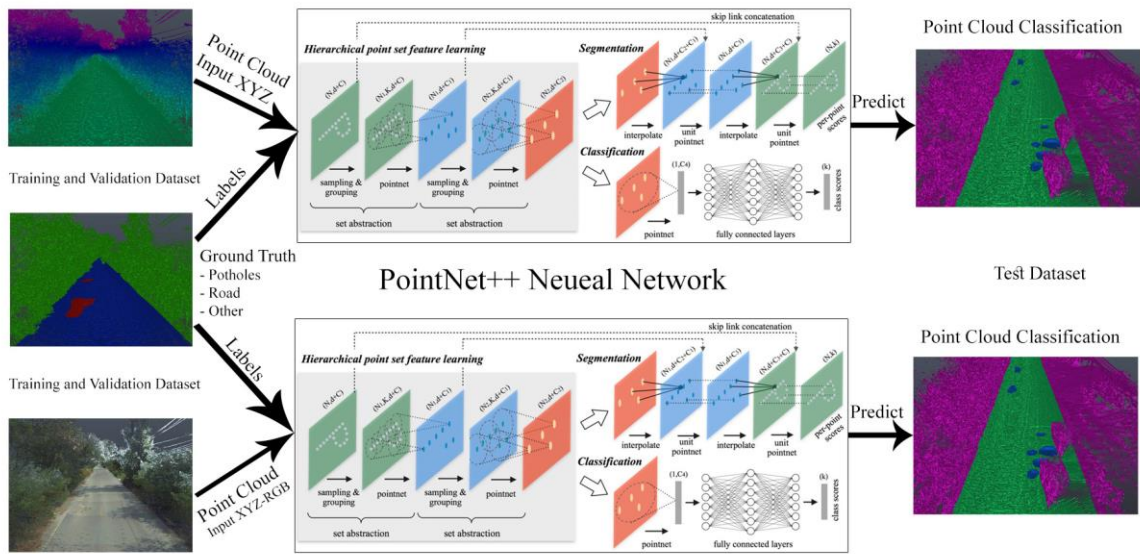


Figure 4: Point cloud data imported into CNN with PointNet++ for training, validating, and testing the classification

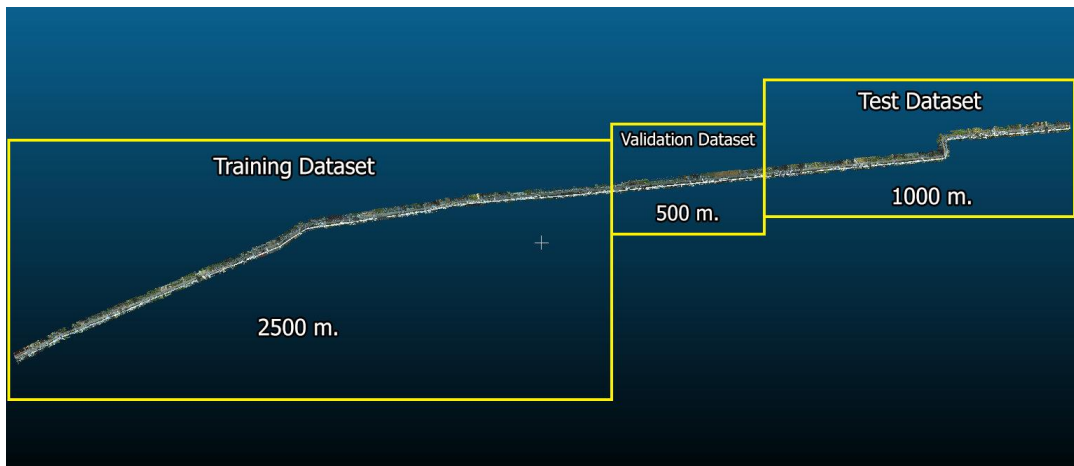


Figure 5: The division of the point cloud data into training, validation, and testing datasets

Hence, it was necessary to use 25% of the data as the sample data for the classification test. In this study, data were divided into three datasets, including training, validation, and testing. Since the point cloud data acquired from MMS were approximately 4,000 meters, they were divided into 2,500 meters of the training dataset, 500 meters of the validation dataset, and 1,000 meters of the testing dataset for an actual classification (Figure 5).

2.6 To Assess for Accuracy, Quality Assessment and Statistical Hypothesis Testing were Conducted

Firstly, the quality of the classification was assessed using the overall accuracy and intersection over union (IoU) values for each class as exhibited in Equations 2 and 3, where TP, TN, FP, and FN refer

to true positive, true negative, false positive, and false negative points, respectively.

$$\text{Overall Accuracy} = \frac{TP+TN}{TP+TN+FP+FN}$$

Equation 2

$$\text{IoU} = \frac{TP}{TP+FP+FN}$$

Equation 3

Results of point cloud classification obtained from both methods through the calculation of Kappa coefficients in the Confusion Matrix were compared to identify differences between the methods [36].

According to Equation 4, the Z-test has the primary hypothesis of $H_0 : (K_1 - K_2) = 0$ (indifferent classification results) and the secondary hypothesis of $H_1 : (K_1 - K_2) \neq 0$ (different classification results). Hence, H_0 would be rejected when $Z \geq Z_{\alpha/2}$ at the 95% confidence level and the critical value of 1.96. In other words, should the Z value falls outside the critical range of -1.96 to 1.96, it means that the results are significantly different at the 95% confidence level.

$$Z = \frac{|K_1 - K_2|}{\sqrt{\text{var}(K_1) - \text{var}(K_2)}} \quad \text{Equation 4}$$

When K_1, K_2 refers to Kappa estimates of Confusion Matrices 1 and 2, and $\text{var}(K_1) - \text{var}(K_2)$ to estimated Kappa variance.

3. Results

3.1 Training

In this study, only 300 epochs were used for the training because the overall accuracy and the mean loss started to remain constant from the 200th epochs onward (Figure 6). The training lasted approximately 30 minutes per epoch. Furthermore, mean loss, overall accuracy, mean IoU, and class IoU were obtained upon completing each epoch. The data were divided into three classes in this training: potholes,

roads, and other objects. The best training results are shown in Table 1, and the results of each epoch are shown in Figure 6. From the results, the proposed method using RGB values from the point cloud for training yielded more effective data classification learning. Although the IoU values obtained from classifying roads and other objects were similar, the mean loss and IoU values from classifying potholes differed significantly. According to the results acquired from the two training methods (Table 1), the most effective training results within the method without RGB values suggested that the IoU of potholes (0.698) was lower than the IoUs of roads (0.961) and other objects (0.989), indicating that the training led to effective learning. Classifying potholes is more complicated than classifying roads and other objects because potholes do not collectively appear with a fixed characteristic and often blend in with the road surface. Therefore, this study introduced the use of RGB color values in training, and the results indicated that the method was more beneficial to the training since the mean loss was reduced from 0.068 to 0.044, and the IoUs of potholes improved from 0.698 to 0.848. Nevertheless, the overall accuracy values, the IoUs of roads, and the IoUs of other objects were similar. Therefore, it is safe to conclude that using RGB values in training increased prediction accuracy, especially in the IoU of potholes.

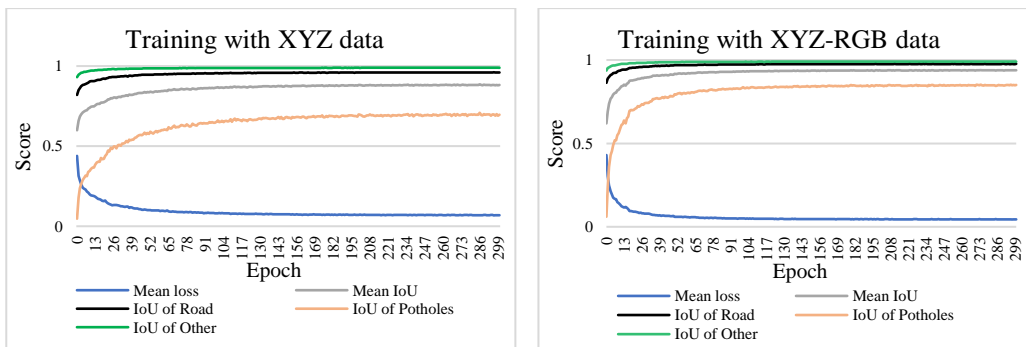


Figure 6: The results of the two training methods using the point cloud data with and without RGB values

Table 1: The results from training CNN with the training dataset

	XYZ	XYZ-RGB
Mean Loss	0.068	0.044
Overall Accuracy	0.989	0.992
Mean IoU	0.883	0.938
IoU of Potholes	0.698	0.848
IoU of Road	0.961	0.976
IoU of Other	0.989	0.990

3.2 Validation

Every five epochs of training with the training dataset, validation was carried out using the validation dataset to store the classification model for the final classification with the testing dataset. Note that the system only stored models with higher test values than their previous epoch. Each validation produced overall accuracy values, mean IoUs, and class IoUs. Table 2 illustrates the highest validation results achieved by the two methods every five-epoch training interval. According to the diagram in Figure 7, the RGB method produced the most accurate prediction in line with the previous training results. Remarkably, the IoU of potholes of the mentioned method (0.813) produced significantly higher results than that of the non-RGB method (0.700), meaning that the RGB method was incredibly and significantly accurate at classifying potholes compared to roads and other objects.

3.3 Classifying the Testing Dataset

To test for learning efficiency, the best performing models obtained from the training and validation were exposed to the testing dataset, a completely new dataset that contains a kilometer of point cloud data and 72 million points approximately. The results acquired from the methods with and without RGB values are illustrated in Figure 8. Overall, the classifications of roads and other objects by both

methods were similarly effective, with the non-RGB method producing more errors when classifying road surfaces by mistaking other objects for road surfaces. Furthermore, the classification of potholes was adequately practical but less effective than that of roads and other objects. The classification was assessed for quality based on the IoU values of each class. According to Table 3, the mean IoU produced by the non-RGB method was 0.826 and the RGB method of 0.858. Although the IoUs of roads and other objects were similar at 0.94 and 0.95, the IoUs of potholes were different. More specifically, the non-RGB method produced a value of 0.595 and the RGB method of 0.667. This discovery is consistent with the results obtained from the training and validating sessions: the method with RGB values outperformed the non-RGB method. Furthermore, when the previously unknown testing dataset was classified, the resulting mean IoUs were 6% and 7.2% lower than those obtained from the validation by the non-RGB and RGB methods, respectively. Furthermore, after assessing the classification quality of both methods through the Confusion Matrix, results revealed that the non-RGB method yielded an overall accuracy of 96.8% and a Kappa coefficient of 0.94 (Table 4). Similarly, the RGB method produced an overall accuracy of 97.5% and a Kappa coefficient of 0.95 (Table 5).

Table 2: Validation results from the validation dataset

	XYZ	XYZ-RGB
Overall Accuracy	0.991	0.994
Mean IoU	0.886	0.930
IoU of Potholes	0.700	0.813
IoU of Road	0.968	0.983
IoU of Other	0.991	0.993

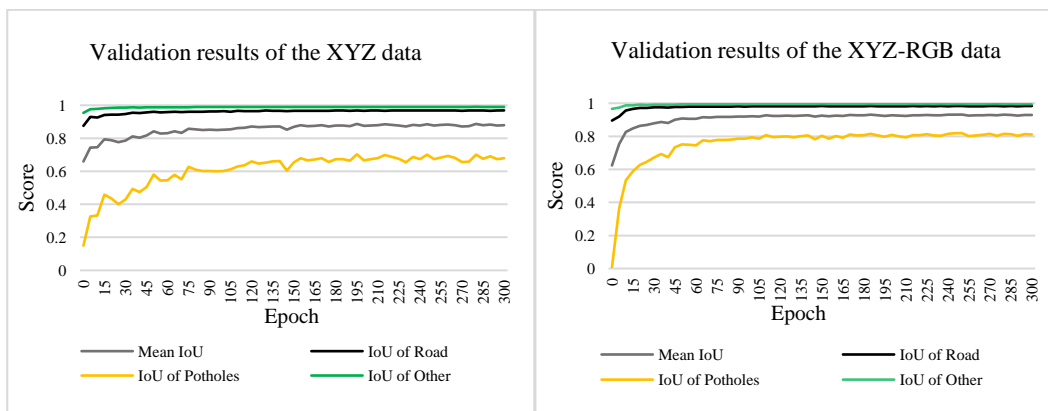


Figure 7: The validation results of the two training methods using the point cloud data with and without RGB values

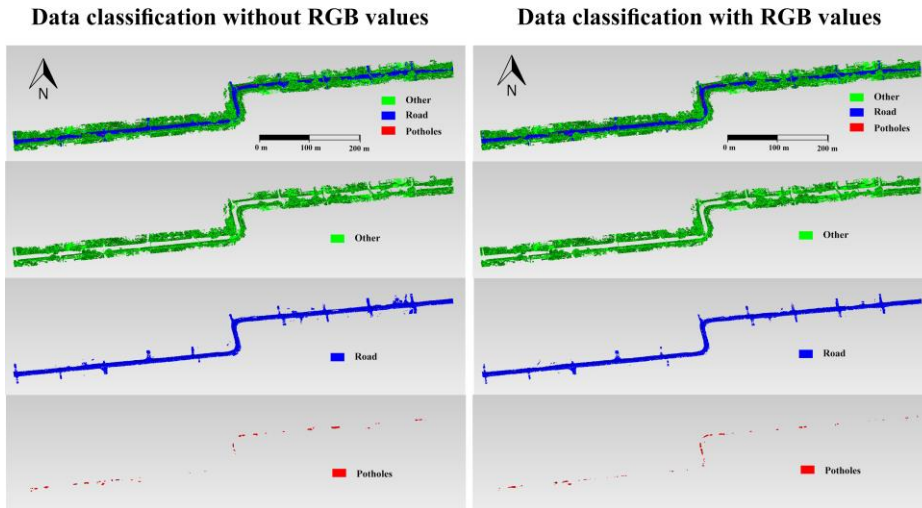


Figure 8: Data classification by both methods

Table 3: The IoU values of both methods

	XYZ	XYZ-RGB
Mean IoU	0.826	0.858
IoU of Potholes	0.595	0.667
IoU of Road	0.942	0.954
IoU of Other	0.941	0.954

Table 4: The Confusion Matrix of the non-RGB method

Point Cloud Classification without RGB					
Classes	Ground Truth			Total	UA, %
	Potholes	Road	Other		
Potholes	510197	164134	24202	698533	73.0
Road	157417	37435563	1673580	39266560	95.3
Other	1533	301056	31676656	31979245	99.0
Total	669147	37900753	33374438	71944338	
PA, %	76.3	98.8	94.9		
Overall Accuracy	96.8%				
Kappa	0.94				

Table 5: The Confusion Matrix of the RGB method

Point Cloud Classification with RGB					
Classes	Ground Truth			Total	UA, %
	Potholes	Road	Other		
Potholes	533176	129340	1342	663858	80.31
Road	135651	37580084	1342928	39058663	96.21
Other	320	191329	32030168	32221817	99.41
Total	669147	37900753	33374438	71944338	
PA, %	79.68	99.15	95.97		
Overall Accuracy	97.50%				
Kappa	0.95				

Table 4 and Table 5 show that the overall accuracy value of the point cloud classification with RGB (97.5%) was slightly higher than that produced by the non-RGB method (96.8%). Consistently with the producer accuracy and user accuracy values obtained from other point cloud classifications, the RGB method yielded increased accuracy. Since the point cloud classifications of roads and other objects typically produce similarly effective accuracy performance, several studies tend to publish similar results. Although the classification of potholes may produce a significantly different result compared to roads and other objects, potholes tend to be regarded as a data class with fewer points to classify. Hence, they tend not to tip the scale of the overall accuracy in broader results.

After the Z-test was administered to the two methods to compare differences in accuracy performance at the 95% confidence level, the Z value of this test was 261.27, falling outside the critical range of -1.96 to 1.96. Furthermore, based on the Z-test and the Confusion Matrix, both methods yielded a significantly different classification outcome, with the RGB method providing a more satisfying classification accuracy with statistical significance when classifying potholes.

4. Discussion

According to Figure 6, the IoUs of potholes from each training epoch improved more slowly in the non-RGB method than in the RGB one, suggesting that the 360-degree panoramic cameras helped enhance the training outcomes. After training the

non-RGB and RGB methods, the best-obtained performance was similar at the overall accuracy values of 98.9% and 99.2%, respectively. However, their IoUs of potholes were 69.8% and 84.8%, respectively. The validation results of each epoch also produced results in a similar fashion compared to the training results, which are lower than the IoUs of roads and other objects. These figures indicate that potholes are a challenging class for CNN to learn and classify. This difficulty could be due to their unique and non-fixed characteristics in the point cloud data since potholes could emerge in any shape on the road surface, confusing the model when classifying.

The final classification with the testing dataset revealed that the non-RGB and RGB methods respectively produced 2.3% and 1.9% decreased overall accuracy values compared to the previous validation classification results. However, the IoUs of roads, potholes, and other objects decreased by 2.6%, 10.5%, and 5.1% for the non-RGB method and 2.9%, 14.6%, and 3.9% for the RGB method, respectively. Evidently, the models obtained from the training and validation were less effective in classifying potholes from previously unknown data than in classifying roads and other objects. However, the Z-test indicated that both methods still produced significantly different outcomes.

Based on the final test with the testing dataset, the non-RGB method was much less accurate and produced several significant errors, such as mistaking other objects for road surfaces. These errors occurred in several locations, especially at road boundary lines and road curbs with a similar level (Figure 9).

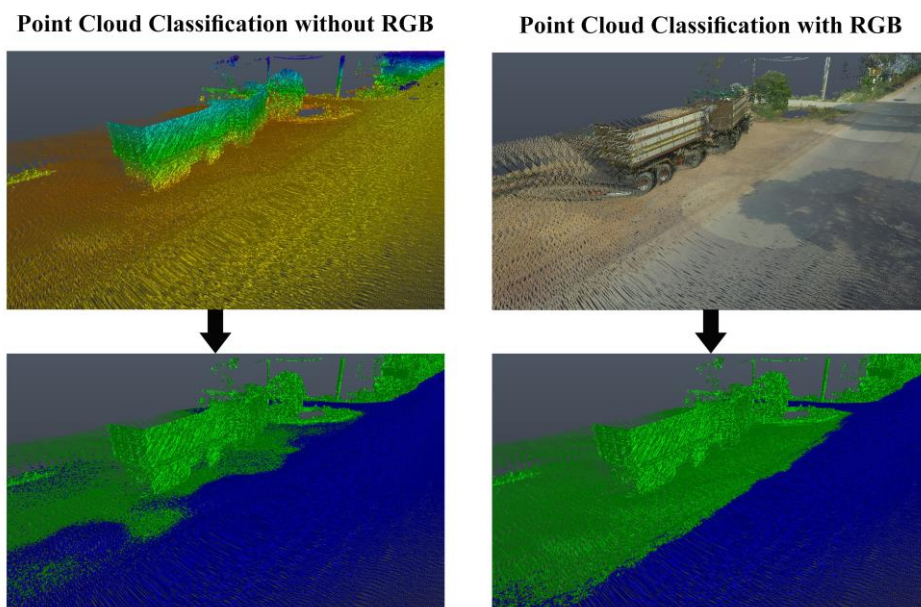


Figure 9: Classifying road boundaries with both methods

These errors are what make the two methods distinctive in classification accuracy. Apparently, the model could learn to classify objects more accurately when seeing RGB color values at road curbs.

When using CNN to classify data, classification efficiency depends on its training dataset. For example, if the sample data contain images of a pothole under the shadow of a tree or building, the model would learn to recognize such a pothole as is and be able to classify a similar pothole in the future. The model is projected to have the capability to classify data by weighing between XYZ and RGB values. Evidently, using CNN to classify potholes directly based on the point cloud data was an exciting endeavor since CNN was able to distinguish objects from 3D point data in combination with color values with a degree of satisfying efficiency. Furthermore, point cloud data can extensively be utilized in the 3D analysis. MMS is point cloud survey equipment that can provide detailed road surface data since its platform can be mounted on vehicles. Nonetheless, there are now other platform options, such as UAV LiDAR, that can also be used to collect 3D data, and they are more affordable than MMS. In addition, UAV LiDAR surveys can provide data with similar quality to MMS, especially when flying at low altitudes. Obtained point cloud data can also be used to classify objects and detect road damage. Therefore, UAV LiDAR seems to be an adequate alternative for further road damage exploration.

5. Conclusion

Based on the point cloud classification performance of the two methods with the three datasets of training, validation, and testing, the method with XYZ and RGB values was more effective than the method with only XYZ values in the PointNet++ architecture. In addition, although both methods did not produce a significantly different outcome when classifying roads and other objects, the RGB method has tremendously outperformed the non-RGB one in pothole classification.

When introducing the training and validation datasets, both methods could learn to classify roads and other objects at a face pace. However, they were slower at learning to classify potholes. The RGB method had a faster classification learning pace during the training and validation. Furthermore, based on the final test with the testing dataset, IoUs of roads and other objects slightly decreased, while those of the potholes decreased considerably. Evidently, classifying potholes from point clouds is a challenging task. Although the RGB method could improve the classification performance, its effectiveness in classifying potholes remained low

compared to other classes, such as roads and other objects.

Based on the results, the classification of potholes became more accurate when using the point cloud data that contained XYZ and RGB values compared to the method that only utilized XYZ values. However, the degree of accuracy achieved in this study was purely based on available data obtained from the survey site. Hence, classification accuracy might change with new datasets of new locations and different pothole characteristics.

In terms of suggestions, Since the road data employed in this study mainly include straight roads with junctions leading to alleys and do not contain primary junctions, such as three-way and four-way, further studies are suggested to utilize point cloud data with primary junctions to assess the effectiveness of road classification at junctions. Furthermore, since the study area contains various types of damage, such as potholes, rutting, and depression. Future studies are recommended to classify road damages according to damage types. In addition, since there were limitations in finding a survey site with consistent and long-distance road surface problems with potholes, further studies are suggested to find areas with lengthy potholes to ensure more pothole samples could be collected and used to train CNN to achieve a higher precision classification performance.

Acknowledgment

The authors would like to express their gratitude to Topcon Instruments (Thailand) Co. Ltd. for providing MMS IP-S3 to collect the field data.

References

- [1] Zakeri, H., Nejad, F. M. and Fahimifar, A., (2017). Image Based Techniques for Crack Detection, Classification and Quantification in Asphalt Pavement: A Review. *Archives of Computational Methods in Engineering*, Vol. 24, 935-977.
- [2] Kim, T. and Ryu, S., (2014). Review and Analysis Of Pothole Detection Methods. *Journal of Emerging Trends in Computing and Information Sciences*, Vol. 5, 603-608.
- [3] Ch, G. L., Sankar, V. U. and Yellampalli, S. S., (2021). Image based Road Distress Detection. *International Conference on Smart Generation Computing, Communication and Networking (SMART GENCON)*.1-6.
- [4] Jo, Y., Ryu, S. K. and Kim, Y. R., (2016). Pothole Detection Based on the Features of Intensity and Motion. *Transportation Research Record*, Vol. 2595, 18-28.

- [5] Wang, P., Hu, Y., Dai, Y. and Tian, M., Asphalt Pavement Pothole Detection and Segmentation Based on Wavelet Energy Field. *Mathematical Problems in Engineering*, Vol. 2017(8),1-13. DOI:10.1155/2017/1604130
- [6] Jindal, A. and Nagarajan, K., (2019). Detection of Potholes and Speedbumps by Monitoring Front Traffic. *SAE Technical Paper 2019-01-5031*, <https://doi.org/10.4271/2019-01-5031>
- [7] Masud, A. K. M. J. A., Sharin, S. T., Shawon, K. F. T. and Zaman, Z., Pothole Detection Using Machine Learning Algorithms. *15th International Conference on Signal Processing and Communication Systems (ICSPCS)*. 1-5, 2021.
- [8] Lakmal, H. K. I. S. and Dissanayake, M. B., (2020). Pothole Detection with Image Segmentation for Advanced Driver Assisted Systems. *IEEE International Women in Engineering (WIE) Conference on Electrical and Computer Engineering (WIECON-ECE)*. 308-311.
- [9] Li, R. and Liu, C., Road Damage Evaluation via Stereo Camera and Deep Learning Neural Network. *IEEE Aerospace Conference (50100)*. 1-7.
- [10] Chen, H., Yao, M. and Gu, Q., (2020). Pothole Detection Using Location-Aware Convolutional Neural Networks. *International Journal of Machine Learning and Cybernetics*, Vol. 11(4), 899-911.
- [11] Darapaneni, N., Reddy, N. S., Urkude, A., Paduri, A. R., Satpute, A. A., Yogi, A., Natesan, D. K., Surve, S. and Srivastava, U., (2021). Pothole Detection Using Advanced Neural Networks. *IEEE 12th Annual Information Technology, Electronics and Mobile Communication Conference (IEMCON)*. 0567-0572.
- [12] Jana, S., Thangam, S., Kishore, A., Kumar, V. S. and Vandana, S., (2022). Transfer Learning Based Deep Convolutional Neural Network Model for Pavement Crack Detection from Images. *International Journal of Nonlinear Analysis and Applications*, Vol. 13(1), 1209-1223.
- [13] Becker, Y. V. F., Siqueira, H. L., Matsubara, E. T., Gonçalves, W. N. and Marcato, J. M., (2019). Asphalt Pothole Detection in UAV Images Using Convolutional Neural Networks. *IGARSS 2019 - 2019 IEEE International Geoscience and Remote Sensing Symposium*. 56-58.
- [14] Pan, Y., Zhang, X., Cervone, G. and Yang, L., (2018). Detection of Asphalt Pavement Potholes and Cracks Based on the Unmanned Aerial Vehicle Multispectral Imagery. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, Col. 11(10), 3701-3712.
- [15] Ravi, R., Habib, A. and Bullock, D., (2020). Pothole Mapping and Patching Quantity Estimates Using Lidar-Based Mobile Mapping Systems. *Transportation Research Record*, Vol. 2674(9), 124-134.
- [16] Zhang, D., Zou, Q., Lin, H., Xu, X., He, L., Gui, R., and Li, Q., (2018). Automatic Pavement Defect Detection Using 3D Laser Profiling Technology. *Automation in Construction*, Vol. 96, 350-365.
- [17] Puente, I., González-Jorge, H., Martínez-Sánchez, J. and Arias, P., (2013). Review of Mobile Mapping and Surveying Technologies. *Measurement*, Vol. 46(7), 2127-2145.
- [18] Ma, L., Li, Y., Li, J., Wang, C., Wang, R. and Chapman, M. A., (2018). Mobile Laser Scanned Point-Clouds for Road Object Detection and Extraction: A Review. *Remote Sensing*, Vol. 10(10). DOI:10.3390/rs10101531.
- [19] Xiang, B., Yao, J., Lu, X., Li, L., Xie, R. and Li, J., (2018). Segmentation-Based Classification for 3D Point Clouds in the Road Environment. *International Journal of Remote Sensing*, Vol. 39, 1-31.
- [20] Yang, M., Liu, X., Jiang, K., Xu, J., Sheng, P. and Yang, D., (2019). Automatic Extraction of Structural and Non-Structural Road Edges from Mobile Laser Scanning Data. *Sensors*, Vol. 19(22). DOI:10.3390/s19225030.
- [21] Wang, D., Wang, J., Scaioni, M. and Si, Q., (2020). Coarse-to-Fine Classification of Road Infrastructure Elements from Mobile Point Clouds Using Symmetric Ensemble Point Network and Euclidean Cluster Extraction. *Sensors*, Vol. 20(1). DOI:10.3390/s20010225.
- [22] De Blasiis, M. R., Di Benedetto, A. and Fiani, M., (2020). Mobile Laser Scanning Data for the Evaluation of Pavement Surface Distress. *Remote Sensing*, Vol. 12(6). <https://doi.org/10.3390/rs12060942>.
- [23] Yang, B., Liu, Y., Dong, Z., Liang, F., Li, B. and Peng, X., (2017). 3D Local Feature BKD to Extract Road Information from Mobile Laser Scanning Point Clouds. *ISPRS Journal of Photogrammetry and Remote Sensing*, Vol. 130, 329-343.
- [24] Serna, A. and Marcotegui, B., (2014). Detection, Segmentation and Classification of 3D Urban Objects using Mathematical Morphology and Supervised Learning. *ISPRS Journal of Photogrammetry and Remote Sensing*, Vol. 93, 243-255.

- [25] Zhongyang, Z., Yinglei, C., Xiaosong, S., Xianxiang, Q. and Li, S., (2018). Classification of LiDAR Point Cloud Based on Multiscale Features and PointNet. *Eighth International Conference on Image Processing Theory, Tools and Applications (IPTA)*. 1-7.
- [26] Zhang, L., Sun, J. and Zheng, Q., (2018). 3D Point Cloud Recognition Based on a Multi-View Convolutional Neural Network. *Sensors (Basel, Switzerland)*, Vol. 18(11). doi: 10.3390/s18113681
- [27] Jing, H. and Suya, Y., (2016). Point Cloud Labeling Using 3D Convolutional Neural Network, *23rd International Conference on Pattern Recognition (ICPR)*. 2670-2675.
- [28] Qi, C., Su, H., Mo, K. and Guibas, L. J., (2017). PointNet: Deep Learning on Point Sets for 3D Classification and Segmentation. *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 77-85.
- [29] Qi, C. R., Yi, L., Su, H. and Guibas, L. J., (2017). PointNet++: Deep Hierarchical Feature Learning On Point Sets In A Metric Space," in *Proceedings of the 31st International Conference on Neural Information Processing Systems*, Long Beach, California, USA, 2017, pp. 5105-5114.
- [30] Lehtomäki, M., Jaakkola, A., Hyypä, J., Lampinen, J., Kaartinen, H., Kukko, A., Puttonen, E. and Hyypä, H., (2016). Object Classification and Recognition from Mobile Laser Scanning Point Clouds in a Road Environment. *IEEE Transactions on Geoscience and Remote Sensing*, Vol. 54(2), 1226-1239.
- [31] Yu, Y., Li, J., Guan, H. and Wang, C., (2015). Automated Extraction of Urban Road Facilities Using Mobile Laser Scanning Data. *IEEE Transactions on Intelligent Transportation Systems*, Vol. 16(4), 2167-2181.
- [32] Ma, L., Li, Y., Li, J., Wang, C., Wang, R. and Chapman, M., Mobile Laser Scanned Point-Clouds for Road Object Detection and Extraction: A Review. *Remote Sensing*, Vol. 10(10).
- [33] Li, J., Cao, S. and Yu, Y., Use of Mobile LiDAR in Road Information Inventory: A Review AU - Guan, Haiyan. *International Journal of Image and Data Fusion*, Vol. 7(3), 219-242.
- [34] Olsen, M. J., Roe, G. V., Glennie, C. L., Persi, F. M., Reedy, M., Hurwitz, D. S., Williams, K., Tuss, H., Squellati, A. and Knodler, M. A., (2013). Guidelines for the Use of Mobile LIDAR in Transportation Applications. *NCHRP Report*, 2013.
- [35] Baheti, P., (2022). Train Test Validation Split: How To & Best Practice. [Online]. Available: <https://www.v7labs.com/blog/train-validation-test-set>. [Accessed: Jul. 10, 2022].
- [36] Congalton, R. G. and Green, K., (2019). *Assessing the Accuracy of Remotely Sensed Data: Principles and Practices*: CRC Press.