# ACCURACY ASSESSMENT OF A SLAM-ACQUIRED POINT CLOUD DATA USING A VARIETY OF CLASSIFICATION APPLICATIONS

H.A. Ingles, A.M.M. Legaspi, C.J.S. Sarmiento, A.R.C. Claridades*

Department of Geodetic Engineering, University of the Philippines Diliman, Quezon City, Philippines – (haingles, amlegaspi2, cssarmiento, acclaridades)@up.edu.ph

**KEY WORDS:** Point Cloud Classification, LiDAR, Accuracy Assessment, SLAM

**ABSTRACT:**

Laser scanning techniques, such as Simultaneous Localization and Mapping (SLAM), produce three-dimensional data representing the real world, which may provide significant information for Building Information Models (BIM). These processes produce 3D point clouds, which require classification before being used in various applications such as structural assessments. However, most widely available software applications for classifying 3D point clouds are proprietary, giving an incomplete depiction of how the data is manipulated and processed. Thus, this research aims to assess the accuracy of the different classification applications in classifying the 3D point cloud data and perform a comparative analysis of the results. Precision, Recall, F1-score, and Accuracy are the evaluation metrics used to assess the classified 3D point cloud data. Results for Precision and Recall show that some of the applications can classify a particular class, the Ground and Building classes. However, the overall performance of the classification method, which is evaluated through the F1-score, produced low values. Results for the F1-score demonstrate that these low values indicate low overall reliability of the classification results despite high values for Accuracy. Based on the conducted experiments, further research is suggested to investigate the effect of increasing dataset size and equalizing class sizes used in classification.

## 1. INTRODUCTION

Building Information Models (BIM) are essential to the digital infrastructure that makes up a Digital Twin (DT). These datasets represent real-world physical structures and are crucial for generating virtual representations of buildings and performing structural analysis or overall lifecycle assessments. In recent years, 3D mapping technologies, such as laser scanning, have been utilized to generate such models. Laser scanning is a procedure where lasers are used to survey and gather measurements of the area or object needed to be scanned to create a 3D map, resulting in point cloud data (Digiscript Philippines, 2021). Point cloud data represent real-life objects and surfaces in a virtual environment (Duan et al., 2019). Such procedures use Light Detection and Ranging (LiDAR), which are sensors that provide precise and accurate measurements of 3D objects (Elhashash et al., 2022).

Simultaneous Localization and Mapping (SLAM) is a mapping process wherein a device simultaneously positions itself in the area where it is mapping. LiDAR-based SLAM produces scanning results that are highly accurate due to the flexibility of the sensor, even without prior knowledge of the site, and it can be used in both indoor and outdoor environments with either light or no light conditions since LiDAR is an active sensor.

Three-dimensional point cloud data undergoes different processes depending on the user's purpose, and one of the processes that this study focused on is the classification of 3D point clouds. This process labels the segmented point cloud data (Grilli et al., 2017). Since BIM is a primarily semantic model, classification is essential to generate data from 3D point clouds. If the data from the classification is inconsistent, the unreliability of BIM data will also increase and potentially create errors in BIM applications.

Available 3D point cloud classification applications may be web-based (Day, 2020) or software-based (Pix4D, 2018). These applications are usually proprietary and do not publicly disclose their algorithms. Hence, users are not usually informed of the detailed methodologies of how classified datasets are generated and, correspondingly, how accurate the results are. With this, this paper aims to assess the accuracy of various point cloud classification applications and perform a comparative analysis based on specific evaluation metrics. The paper is organized as follows. The next section briefly discusses relevant studies, while the following section discusses the study's methodology. Then, the fourth section contains the results of the accuracy assessment. Finally, the last section summarizes and discusses future directions for this study.

## 2. RELATED LITERATURE

BIM is essential in analyzing smart city infrastructures (MgBere et al., 2018). BIM is a technology that can represent 3D information models or 3D maps of an infrastructure project, aiding stakeholders in planning, designing, constructing, operating, and managing a facility in a timely and cost-effective manner. (Goyal et al., 2020). Point cloud data from LiDAR and classification and segmentation algorithms are used to generate such models to create reliable 3D virtual environments (Jaren & Arranz, 2021).

Simultaneous Localization and Mapping (SLAM) is a method becoming more accepted in the field of mapping, as its localization feature has been vital in data gathering and positioning. The process SLAM uses is identifying an unfamiliar environment and localizing while referencing through a map. SLAM's objective is to update the position of the equipment while understanding its environment (Kuzmin, 2018; Sossalla et al., 2021).

Three-dimensional point cloud data, which may be collected through SLAM, are prone to noise due to imperfections of the technology and data acquisition; denoising is one of the methods used to clean the 3D point cloud before further processing (Cattai et al., 2022). Additionally, there are noise points not at the

* Corresponding author


https://doi.org/10.5194/isprs-archives-XLVIII-4-W8-2023-307-2024 |  

scanned surface called outliers (Landa, 2013). One data cleansing process applied for as-built modeling is through outlier removal, filling holes and gaps, and balancing its density (Rashidi and Brilakis, 2016).

In point cloud data processing, segmentation and classification are the most essential processes since they extract the information needed for a specific application (Nygren and Jasinski., 2016). Segmentation is the process of identifying and grouping points into certain regions with the same attributes, while classification or semantic segmentation is the process of labeling the segmented data (Grilli et al., 2017). Classifying point clouds is considered one of the most helpful processes in performing structural analysis on buildings (Ntiyakunze & Inoue, 2023). As buildings deform over time, overall life cycle assessments are crucial, especially with hazard assessments, and this could quickly be done through semantic segmentation (Zahs, 2023). Aside from this application, proper model creation could also be generated through point cloud classification (Truong-Hong et al., 2021).

A study about a survey on deep learning-based segmentation, detection, and classification for 3D point clouds shows that one approach to 3D point cloud classification is deep learning. It is a classification type involving deep neural networks that assess and classify the points in a 3D space. This study surveyed the different deep learning classifications to evaluate the efficiency of these methods in creating a classification model. These models were created through training classification, where strips of 3D point clouds are used to gather information about a specific class and create a model. These models are then used to automatically classify raw data of a 3D point cloud (VinodKumar, 2023).

Statistical metrics of precision and recall can be used to determine the accuracy of the classification process of the 3D point cloud data. Precision is the percentage of correctly identifying the classified points, while recall is the proportion of reference points that were classified correctly (Wang L. et al., 2020). In a similar study about the supervised classification of point cloud data, the point cloud classification was assessed using precision, recall, overall accuracy, and F1-score. F1-score is the harmonic mean function of precision and recall, while overall accuracy is the percentage of correctly classified points. Results showed that different algorithms for classifying point cloud data produce different precision, recall rate, F1-score, and overall accuracy. It is important to remember that choosing the proper method and parameters will significantly impact the classification's accuracy (Atik et al., 2021). The resulting F1-score is observed to be good if it is more than 50% and poor if it is not (Allwright, 2022a). For accuracy, the considered value deemed as good is to have more than 60%. Anything lower than that is determined to have poor accuracy (Allwright, 2022b). These thresholds were generalized values for F1-score and accuracy as a whole since there are no studies yet on what accuracy fits for structural assessment for BIM.

In the accuracy assessment of 3D point cloud classification, it is essential to note that the overall accuracy is not enough to evaluate the performance of a classification method since it only gives the overall correctness of the classification method. In a study about training data minimization for 3D point cloud classification, their data shows that their overall accuracies are higher compared to the F1-scores. Even though their overall accuracy and F1-score are both acceptable, their value is still different (Morsy & Shaker, 2022). The difference in value is because accuracy and F1-score evaluate the classification model differently. F1-score considers how the data is distributed and is often used when the data is imbalanced. It also looks at false negatives critically compared to accuracy. For accuracy, it does not consider how the data is distributed, and it is often used when the data is balanced. Both accuracy and F1-score should exist in assessing the accuracy and performance of a classification method because they provide complementary insights into different aspects of the method's effectiveness (Zach, 2021).

## 3. METHODOLOGY

Figure 1 illustrates the overall methodology of this study. We begin by collecting and denoising a SLAM-based LiDAR data in the study area and proceed with the classification using various point cloud classification software programs. We then analyze the accuracy of the results and compare the resulting accuracy metrics for each program. The following sections detail each step in this methodology.

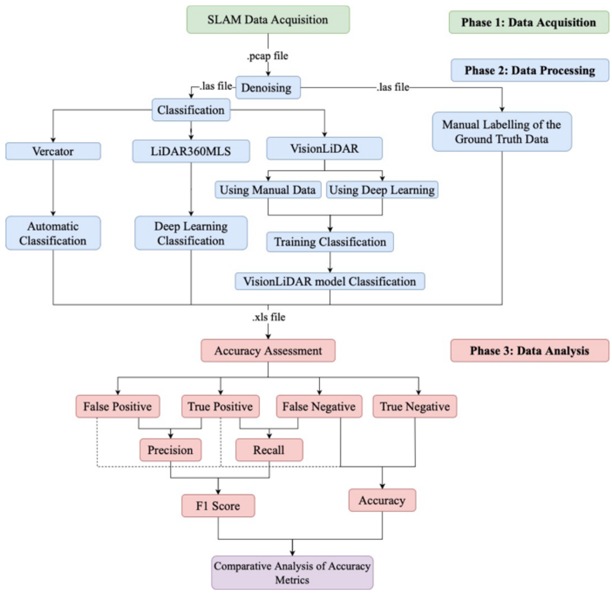

**Figure 1.** Methodological Framework

### 3.1 Study Area and Data Collection

The study area chosen is the National Institute of Molecular Biology and Biotechnology (NIMBB), a building in the University of the Philippines, Diliman. NIMBB provides an optimal site for the research's scanning process since the infrastructure does not have too many obstructions around it, making the objects easily distinguishable. The Foxtech SLAM 100, a handheld mobile LiDAR laser scanner, was utilized in scanning the area.

This study's scanning route will follow the existing pathway surrounding the study area. Figure 2 shows the path of the scanning proper. It was ensured that the route was in a loop closure and that there was a 15-meter overlap from the starting to the ending mark. The green arrows represent the direction followed during the scan, and the dashed blue lines represent the overlap done. During the scanning proper, there must be little to no moving entity in the scanning area, so there will be less noise in the gathered point cloud data. Moreover, before moving around the study area, the equipment requires a 60-second initialization time to scan the starting point of the scanning proper.

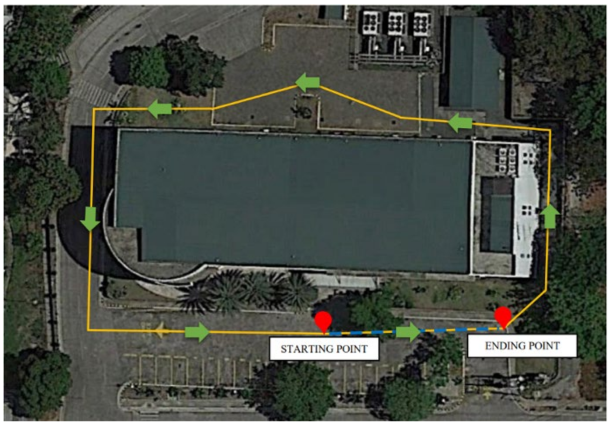


**Figure 2.** Scanning Route around the Study Area

### 3.2 Data Processing

The scanned data were denoised in SlamGo Post Pro Software, the included application of the equipment. After denoising, the ground truth data was manually by annotating the surfaces of the 3D point cloud. One process of manually labeling 3D point clouds is through the use of software that contains annotation tools that make use of volumetric shapes like cuboids, cylinders, spheres, and other free-hand annotation tools; data extracted from these annotation tools are Class name, Class ID, Instance ID, and several points (Ibrahim et al., 2021). In the study, the researchers used CloudCompare software to manually label the 3D point cloud to obtain the ground truth data.

Table 1 shows the different classification applications utilized and their access features. The classification model employed follows the ASPRS standard for point cloud classification, although there are some added features to some of the classification models. Additionally, deep learning classification was utilized for most applications except for VisionLiDAR. In Vision LiDAR, training classification was first done to create a model for the deep learning classification of the 3D point cloud.

| Application | Application Type | Application Access | Application Data Limitation | Classification Model | Classification Utilized |
|---|---|---|---|---|---|
| Vercator | Web-Based | Free Trial | >100,000 kb | Highway | Deep Learning |
| LiDAR360MLS | Software-Based | Free Trial | >100,000 kb | GVPointCloud | Deep Learning |
| Vision LiDAR | Software-Based | Free Trial | >100,000 kb | GeoPlus MobileGeneral | Training and Deep Learning |

**Table 1.** Classification Applications Overview

The automatic classification was done in Vercator, a web-based proprietary application. The process done by the application was to segment the 3D point cloud first, which pertains to grouping up points with the same attributes. Classification for deep learning was done in LiDAR360MLS, a software-based application. There were two processes done using VisionLiDAR. This first process used the manually classified data for training to create a classification model. The second process used a pre-trained model already in the application for training. This generated a classification model based on the researcher's gathered data. In the training process using the manually classified data, the researchers divided the data into thirds, where one-third was used for validation, and two-thirds were used for training.

### 3.3 Accuracy Assessment

A confusion matrix was used in the study to evaluate the effectiveness of the classification process for both the automated classified data and the ground truth data (actual value). In a classification model from a web-based and software-based application, the 3D point cloud data is subjected to different class predictions, which can be analyzed through the confusion matrix. The retrieved elements are the data predicted by the program of the application, and they contain points marked as True Positive (TP) (which are the correctly predicted classified data) and points marked as False Positive (FP) (which are the incorrectly predicted classified data). Outside the retrieved True Positive (TP) elements are the relevant data, also called False Negatives (FN). These are the true classified data that have been excluded from prediction. Lastly, True Negatives (TN) are correctly identified false classified data excluded from the prediction (Poux et al., 2020). The results of the confusion matrix were further processed to assess the accuracy and performance of the classification process.

Elements of the confusion matrix are used to calculate the evaluation metrics, Precision, Recall, F1-score, and Accuracy. These metrics are essential because they provide complementary insights into the effectiveness of the classification method of each application (Zach, 2021). These metrics are illustrated in the following equations.

$$Precision = \frac{TP}{TP+FP} \quad (1)$$

$$Recall = \frac{TP}{TP+FN} \quad (2)$$

$$F1-score = 2 \times \frac{Precision \times Recall}{Precision + Recall} \quad (3)$$

$$Accuracy = \frac{TN + TP}{TN + FN + TP + FP} \quad (4)$$

## 4. RESULTS AND DISCUSSION

Figure 3 shows the manually classified data, which was the basis for assessing the accuracy and performance of each classification model per application. The colors represent the different classes the researchers assigned to each part of the 3D point cloud. These classes were combined classes based on the ASPRS standard. To have an overview of how the researchers set these classes, here are the following definitions for each class is summarized in Table 2.

| ASPRS Point Cloud Class | Description |
|---|---|
| Vehicles | Cars and Motorcycles |
| Vegetation | Medium and High Vegetation |
| Unclassified | Anything that is not within the study area, along with the human and vehicle movements in the 3D point cloud data |
| Building | The main building of NIMBB |
| Manmade | Fences, Street light poles, Signages, Utility poles, and Electrical wires |
| Ground | Low vegetation, Road surface, and Soil ground |

**Table 2.** Description of the classes used for classification

Since the different applications for classification have produced different classes for its classification, it was standardized based on the set classes of the researchers. All standardization processes were done manually in CloudCompare using its annotation tools.

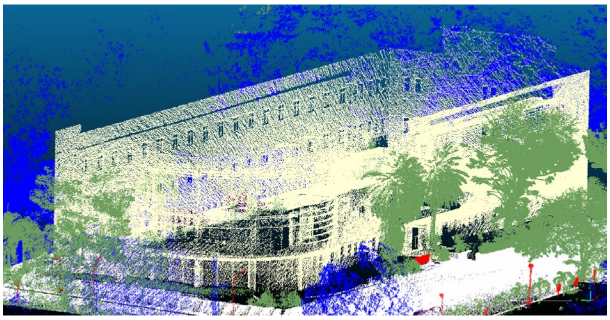

**Figure 3.** Manually Classified 3D Point Cloud

Table 3 shows the confusion matrix formed from the automatic classification of Vercator. The values of the confusion matrix are the number of points of the correctly and incorrectly classified data for a particular class. Additionally, the zero values in the table mean that the class has not been subjected to incorrect classification. Table 4 shows the different accuracy metrics used to evaluate the performance and accuracy of the automatic classification of Vercator. The red and green highlights specify if the obtained data passed the threshold of 75% for each accuracy metric. The thresholds set for each accuracy metric mean that the classification method performed well in the data prediction of classes.

| | VERCATOR | | | | | | |
|---|---|---|---|---|---|---|---|
| | | Building | Vegetation | Ground | Manmade | Vehicle | Unclassified |
| GROUND TRUTH | Building | 483 | 0 | 0 | 0 | 0 | 194 |
| | Vegetation | 24794 | 2736740 | 17674 | 4596 | 2 | 1225281 |
| | Ground | 84603 | 11603 | 3149255 | 106 | 0 | 746617 |
| | Manmade | 1795166 | 436658 | 18605 | 268790 | 0 | 444543 |
| | Vehicle | 651 | 0 | 20 | 0 | 33206 | 17353 |
| | Unclassified | 5 | 0 | 0 | 0 | 0 | 1935 |

**Table 3.** Confusion matrix for Vercator

| | Precision | Recall | F1-score | Accuracy |
|---|---|---|---|---|
| **Building** | 0.0253% | 71.3442% | 0.0507% | 82.7077% |
| **Vegetation** | 85.9259% | 68.2634% | 76.0830% | 84.3849% |
| **Ground** | 98.8605% | 78.8855% | 87.7506% | 92.0207% |
| **Manmade** | 98.2808% | 9.0692% | 16.6060% | 75.4996% |
| **Vehicle** | 99.9940% | 64.8175% | 78.6518% | 99.8364% |
| **Unclassified** | 0.0794% | 99.7423% | 0.1587% | 77.9107% |
| **Average** | **63.8610%** | **65.3537%** | **43.2168%** | **85.3933%** |

**Table 4.** Accuracy assessment computation for Vercator

Based on the precision, the classification model of Vercator does not perform well in predicting which classes are relevant to the ground truth for Building and Unclassified (highlighted in red), resulting in an average precision of 63.8610%. For the recall, the classification model of Vercator does not perform well in predicting the Building, Vegetation, Manmade, and Vehicle classes (highlighted in red), resulting in an average recall of 65.3537%. Now, for the F1-score, the classification model of Vercator has the best performance in predicting the data of the Vegetation, Ground, and Vehicle classes, as highlighted in green in Table 3. Due to the low performances in F1-score, it obtained an average of 43.2168%. Lastly, since the application predicted all classes well, the accuracy of prediction of these classes is good (highlighted in green), with an average accuracy value of 85.3933%.

The confusion matrix for Table 5 contains the data obtained from the deep learning classification of LiDAR360MLS, while the same accuracy metrics are used in Table 6 are employed to assess its accuracy.

| | LIDAR360 | | | | | | |
|---|---|---|---|---|---|---|---|
| | | Building | Vegetation | Ground | Manmade | Vehicle | Unclassified |
| GROUND TRUTH | Building | 1812725 | 61120 | 0 | 0 | 0 | 191414 |
| | Vegetation | 21498 | 3026012 | 45864 | 189005 | 37984 | 1401290 |
| | Ground | 38764 | 78772 | 3106456 | 73700 | 11342 | 829718 |
| | Manmade | 0 | 1037 | 0 | 9825 | 0 | 1302 |
| | Vehicle | 4427 | 0 | 0 | 0 | 9457 | 33339 |
| | Unclassified | 802 | 1118 | 7 | 11043 | 1471 | 19329 |

**Table 5.** Confusion matrix for LiDAR360MLS

| | Precision | Recall | F1-score | Accuracy |
|---|---|---|---|---|
| **Building** | 96.5131% | 87.7723% | 91.9354% | 97.1138% |
| **Vegetation** | 95.5163% | 64.0880% | 76.7078% | 83.3223% |
| **Ground** | 98.5449% | 75.0578% | 85.2125% | 90.2152% |
| **Manmade** | 3.4647% | 80.7711% | 6.6444% | 97.4944% |
| **Vehicle** | 15.6952% | 20.0263% | 17.5982% | 99.1963% |
| **Unclassified** | 0.7805% | 57.2372% | 1.5401% | 77.5702% |
| **Average** | **51.7525%** | **64.1588%** | **46.6064%** | **90.8187%** |

**Table 6.** Accuracy assessment computation for LiDAR360MLS

As shown in Table 4, values of the precision metric is not performing well in predicting the classes relevant to the ground truth for Manmade, Vehicle, and Unclassified (highlighted in red), resulting in an average precision of 51.7525%. Its recall data does not accurately predict the classes of Vegetation, Ground, Vehicle, and Unclassified classes (highlighted in green), resulting in an average recall of 64.1588% (Xiang et al., 2017). Now, for the F1-score, the classification model of LiDAR360MLS performed well in predicting the data of the Building, Vegetation, and Ground classes. Due to the low performances in F1-score, it obtained an average of 46.6064%. Lastly, the accuracy of classification for the LiDAR360MLS is also considered good with an average value of 90.8187%.

Shown in Table 7 is the confusion matrix formed from the classification generated by VisionLiDAR using the manually classified data for training, and different accuracy metrics used to evaluate the performance of the classification done by VisionLiDAR using the manually classified data for training are seen in Table 8.

| | VISIONLIDAR Manual | | | | | | |
|---|---|---|---|---|---|---|---|
| | | Building | Vegetation | Ground | Manmade | Vehicle | Unclassified |
| GROUND TRUTH | Building | 1811779 | 208086 | 8233 | 269176 | 21251 | 326439 |
| | Vegetation | 4968 | 2907876 | 0 | 9915 | 7934 | 1381079 |
| | Ground | 68612 | 42868 | 3072134 | 4 | 7287 | 786841 |
| | Manmade | 5518 | 26 | 0 | 258 | 0 | 811 |
| | Vehicle | 251 | 7000 | 45 | 0 | 24019 | 42791 |
| | Unclassified | 0 | 1531 | 0 | 1271 | 0 | 877 |

**Table 7.** Confusion matrix for VisionLiDAR (Model trained using manually classified data)

| | Precision | Recall | F1-score | Accuracy |
|---|---|---|---|---|
| **Building** | 95.8041% | 68.4992% | 79.8828% | 91.7185% |
| **Vegetation** | 91.8068% | 67.4404% | 77.7594% | 84.9040% |
| **Ground** | 99.7313% | 77.2330% | 87.0520% | 91.7061% |
| **Manmade** | 0.0919% | 3.9014% | 0.1796% | 97.3979% |
| **Vehicle** | 39.7067% | 32.4117% | 35.6902% | 99.2144% |
| **Unclassified** | 0.0345% | 23.8380% | 0.0690% | 76.9417% |
| **Average** | **54.5292%** | **45.5540%** | **46.7722%** | **90.3138%** |

**Table 8.** Accuracy assessment computation for VisionLiDAR (Model trained using manually classified data)

The precision metric shows that the classification by VisionLiDAR did not perform well in predicting which classes, highlighted in red, are relevant to the ground truth for Manmade, Vehicle, and Unclassified classes, resulting in a low average precision of 54.5292%. Correspondingly, the recall metric resulted in an average of 45.5540% which indicates that the classification model produced classes that were not relevant to the ground truth. Only one class, Ground, was in the range of the good recall rate and is highlighted in green. Nevertheless, the average recall was not considered to be in the range of an acceptable recall rate. The average F1-score is 46.7722% coming from the unsatisfactory performance of precision and recall in predicting some of the classes, specifically, the classes highlighted in red which are Manmade, Vehicle, and Unclassified. This F1-score result is not considered good. However, the average accuracy is 90.3138%, indicating that there is a high accuracy and is considered acceptable in the classification of VisionLiDAR using the manually classified data for training.

The confusion matrix from the classification model formed by VisionLiDAR using a pre-trained model from the classification application can be seen in Table 9 and seen in Table 10 are the different accuracy metrics used to assess the classification done in the VisionLiDAR pre-trained model for training.

| | VISIONLIDAR360 Pre-Trained | | | | | | |
|---|---|---|---|---|---|---|---|
| | | **Building** | **Vegetation** | **Ground** | **Manmade** | **Vehicle** | **Unclassified** |
| GROUND TRUTH | **Building** | 1801000 | 232029 | 0 | 268919 | 22381 | 320635 |
| | **Vegetation** | 4949 | 2942241 | 0 | 11177 | 7978 | 1345427 |
| | **Ground** | 68663 | 14076 | 3118837 | 0 | 9845 | 766325 |
| | **Manmade** | 5518 | 22 | 0 | 252 | 0 | 821 |
| | **Vehicle** | 251 | 6997 | 49 | 0 | 26226 | 40583 |
| | **Unclassified** | 0 | 1546 | 0 | 1302 | 0 | 831 |

**Table 9.** Confusion matrix for VisionLiDAR (Pre-Trained Classification Model)

| | Precision | Recall | F1-score | Accuracy |
|---|---|---|---|---|
| **Building** | 95.7785% | 68.0917% | 79.5961% | 91.6203% |
| **Vegetation** | 92.0339% | 68.2374% | 78.3690% | 85.2598% |
| **Ground** | 99.9984% | 78.4071% | 87.8963% | 92.2047% |
| **Manmade** | 0.0895% | 3.8107% | 0.1748% | 97.3885% |
| **Vehicle** | 39.4792% | 35.3898% | 37.3228% | 99.2006% |
| **Unclassified** | 0.0336% | 22.5877% | 0.0671% | 77.5237% |
| **Average** | **54.5688%** | **46.0874%** | **47.2377%** | **90.5329%** |

**Table 10.** Accuracy assessment computation for VisionLiDAR (Pre-Trained Classification Model)

The precision metric showed that VisionLiDAR did not perform well in predicting which classes are relevant to the ground truth for the Manmade, Vehicle, and Unclassified classes. These classes are all highlighted in red. Using the installed model for training in VisionLiDAR still had an average precision of 54.5688%. The recall metric also showed that VisionLiDAR did not perform well, averaging 46.0874%. The result indicates that the classification model produced classes irrelevant to the ground truth. The average F1-score is 47.2377% and is considered poor. This result is due to the unsatisfactory performance of precision and recall in predicting some of the classes. However, the average accuracy is 90.5329%, comparable to previous platforms' results.

Table 11 summarizes the resulting values of the metrics for the different classification applications used in this study. The precision and recall values of the classification model for Vercator, Lidar360MLS, VisionLIDAR manual, and pretrained classification produced values lower than 77.5%, which is an indicator that the classification method did not perform well in predicting which classes are relevant to the ground truth and in predicting the correct point clouds for each class. Since F1-score is the harmonic mean between precision and recall, the average values were also considered a poor score (Allwright, 2022a). Although the average accuracy of each classification application was deemed high, the F1-score proves it is a good metric for checking the overall performance of the classification model of Vercator, especially if the data is imbalanced.

| Classification Application | Precision | Recall | F1-score | Accuracy |
|---|---|---|---|---|
| Vercator | 63.8610% | 65.3537% | 43.2168% | 85.3933% |
| Lidar360MLS | 51.7525% | 64.1588% | 46.6064% | 90.8187% |
| VisionLIDAR (manual classification) | 54.5292% | 45.5540% | 46.7722% | 90.3138% |
| VisionLIDAR (pre-trained classification) | 54.5688% | 46.0874% | 47.2377% | 90.5329% |

**Table 11.** Summary of accuracy assessment

Taking the case of the LiDAR360MLS application, which produced values that are lower than 77.5% for precision and recall, which means that the classification method did not perform well in predicting which classes are relevant to the ground truth and in predicting the correct point clouds for each class. Its average F1-score is 46.6064%, which is considered a poor score for classification performance. However, the average accuracy of 90.8187% demonstrates a good level of accuracy in correctly predicting the classes. This further proves that relying solely on accuracy as a classification metric may not be entirely reliable. (Allwright, 2022b).

Since most of the precision and recall values are low, the performances of the different classification models did not correctly predict the classes of the 3D point cloud. These low results produced poor F1-score values (Allwright, 2022a; Xiang et al., 2017). These results imply that the different classification models have their strengths and weaknesses, as seen in how they predicted the classifications of the researcher's 3D point cloud data. Vercator was able to efficiently assess the Vehicle class while for LiDAR360MLS, the Building class. Both the classification model based on the manually classified data and the pretrained classification model in VisionLiDAR assessed the Ground class effectively. These classes were the ones that had all the accuracy metrics highlighted in green from the tables, indicating the adequate classification done by the applications.

In contrast with the other evaluation metrics, the accuracy of the classification applications is relatively high. This discrepancy between the accuracy and F1-score proves the need for both in evaluating the classification application. This contrast in results has been evident since some classes perform better than others on the various classification applications. Accuracy does not consider the false negative values and the data distribution, which the F1-score does (Zach, 2021).

## REFERENCES

Allwright, S., 2022a. https://stephenallwright.com/good-accuracy-score/.

Allwright, S., 2022b. What is a good F1 score and how do I interpret it? Stephen Allwright. https://stephenallwright.com/good-f1-score/.

Atik, M. E., Duran, Z., Seker, D. Z., 2021. Machine learning-based supervised classification of point clouds using multiscale geometric features. *ISPRS International Journal of Geo-Information*, 10(3). https://doi.org/10.3390/ijgi10030187.

Cattai, T., Delfino, A., Scarano, G., Colonnese, S., 2022. VIPDA: A Visually Driven Point Cloud Denoising Algorithm Based on Anisotropic Point Cloud Filtering. *Frontiers in Signal Processing*, 2. https://doi.org/10.3389/frsip.2022.842570.

Day, M., 2021. Vercator - point clouds in the cloud. AEC Magazine. https://aecmag.com/technology/vercator-point-clouds-in-the-cloud/?fbclid=IwAR0VdFzO0yjfcrp44OYrQLewKaP1QPwZm-3v4vOiAEYpvMA7KYqlgUTfa38.

Digiscript Ph., n.d. 3D Laser Scanning Philippines. Index. Retrieved October 20, 2022, from https://www.digiscriptinc.com/3d-laser-scanning.

Duan, Y., Yang, C., Chen, H., Yan, W., Li, H., 2021. Low-complexity point cloud denoising for LiDAR by PCA-based dimension reduction. *Optics Communications*, 482, 126567. https://doi.org/10.1016/J.OPTCOM.2020.126567.

Elhashash, M., Albanwan, H., Qin, R., 2022. A Review of Mobile Mapping Systems: From Sensors to Applications. *In Sensors* (Vol. 22, Issue 11). MDPI. https://doi.org/10.3390/s22114262.

Goyal, L. K., Chauhan, R., Kumar, R., Rai, H. S., 2020. Use of BIM in Development of Smart Cities: A Review. *IOP Conference Series: Materials Science and Engineering*, 955(1). https://doi.org/10.1088/1757-899X/955/1/012010.

Grilli, E., Menna, F., Remondino, F., 2017. A Review of Point Clouds Segmentation and Classification Algorithms. International Archives of the Photogrammetry, *Remote Sensing and Spatial Information Sciences - ISPRS Archives*, 42(2W3), 339–344. doi.org/10.5194/isprs-archives-XLII-2-W3-339-2017.

Ibrahim, M., Akhtar, N., Wise, M., Mian, A., 2021. Annotation Tool and Urban Dataset for 3D Point Cloud Semantic Segmentation. *IEEE Access*, 9, 35984–35996. https://doi.org/10.1109/ACCESS.2021.3062547.

Kuzmin, M., 2018: Review, Classification and Comparison of the Existing SLAM Methods for Groups of Robots.

Landa, J., Prochazka, D., Štastny, J., 2013. Point cloud processing for smart systems. *Acta Universitatis Agriculturae et Silviculturae Mendelianae Brunensis*, 61(7), 2415–2421. https://doi.org/10.11118/actaun201361072415.

Mgbere, C., Knyshenko, V. A., Bakirova, A. B., 2018. Building information modeling. A management tool for Smart City. *International Scientific and Technical Conference on Computer Sciences and Information Technologies*, 2, 177–182. https://doi.org/10.1109/STC-CSIT.2018.8526634.

Morsy S., Shaker A., 2022. Evaluation of LiDAR-Derived Features Relevance and Training Data Minimization for 3D Point Cloud Classification. https://doi.org/10.3390/rs14235934.

Nygren, P., & Jasinski, M., n.d. A Comparative Study of Segmentation and Classification Methods for 3D Point Clouds.

Ntiyakunze, J., & Inoue, T., 2023. Segmentation of Structural Elements from 3D Point Cloud Using Spatial Dependencies for Sustainability Studies. *Sensors*, 23(4), 1924. doi.org/10.3390/s23041924.

Pix4d, 2018. How to generate the point cloud classification. https://support.pix4d.com/hc/en-us/articles/115004864586-How-to-generate-the-point-cloud-classification?fbclid=IwAR2jnUNjGeE5KBdkGkE1-LN5m1FNlbecmzxymWTNTmKEv94LOqk4QYDMEhs.

Poux, F., Mattes, C., Kobbelt, L., 2020. Unsupervised Segmentation of Indoor 3D Point Cloud: Application to Object-Based Classification. *International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences - ISPRS Archives*, 44(4/W1), 111–118. https://doi.org/10.5194/isprs-archives-XLIV-4-W1-2020-111-2020.

Rashidi, A., Brilakis, I., 2016. Point Cloud Data Cleaning and Refining for 3D As-Built Modeling of Built Infrastructure. *Construction Research Congress 2016: Old and New Construction Technologies Converge in Historic San Juan - Proceedings of the 2016 Construction Research Congress, CRC 2016*, 919–929. https://doi.org/10.1061/9780784479827.093.

Sossalla, P., Rischke, J., Hofer, J., Fitzek, F. H. P., 2021. Evaluating the Advantages of Remote SLAM on an Edge Cloud. *IEEE International Conference on Emerging Technologies and Factory Automation, ETFA*, 2021-September. https://doi.org/10.1109/ETFA45728.2021.9613415.

Truong-Hong, L., Lindenbergh, R., & Nguyen, T. A., 2021. Structural Assessment Using Terrestrial Laser Scanning Point Clouds. *International Journal of Building Pathology and Adaptation*, 40(3), 345–379. doi.org/10.1108/ijbpa-04-2021-0051.

Vinodkumar P., Karabulut D., Avots E., Ozcinar C., Anbarjafari G., 2023. A Survey on Deep Learning Based Segmentation, Detection and Classification for 3D Point Clouds. https://doi.org/10.3390/e25040635.

Wang, L., Huang, Y., Shan, J., He, L., 2018. MSNet: Multi-scale convolutional network for point cloud classification. *Remote Sensing*, 10(4). https://doi.org/10.3390/rs10040612.

Xiang, B., Yao, J., Lu, X., Li, L., Xie, R., Li, J. 2017. Segmentation-Based Classification for 3D Point Clouds in the Road Environment. *Remote Sens*, pp. 1–27. doi.org/10.1080/01431161.2018.1455235.

Zach, 2021. F1 score vs. accuracy: Which should you use? Statology. https://www.statology.org/f1-score-vs-accuracy/.

Zahs, V., Anders, K., Kohns, J., Stark, A., Höfle, B., 2023. Classification of structural building damage grades from multi-temporal photogrammetric point clouds using a machine learning model trained on virtual laser scanning data. *ISPRS J. Photogram. Remote Sens*, 1–29. doi.org/10.48550/arXiv.2302.1.