Classification of 3D Sonar Point Clouds derived Underwater using Machine and Deep Learning (CANUPO and RandLA-Net) Approaches

Simiso Ntuli[1], Mulemwa Akombelwa[1], Angus Forbes[1], Mayshree Singh[2]

[1]Discipline of Land Surveying, School of Engineering, University of KwaZulu-Natal, Durban, South Africa, 215017855@stu.ukzn.ac.za
[2]Wits Mining Institute, Faculty of Engineering and the Built Environment, University of Witwatersrand, Johannesburg, South Africa, mayshree.singh@wits.ac.za

DOI: https://dx.doi.org/10.4314/sajg.v13i2.4

Abstract

The techniques of point cloud classification in aquatic environments have various applications such as landslide hazard mapping, recovery of lost objects, underwater infrastructure inspection, exploration of mineral resources on the seabed, underwater cultural heritage documentation, environmental preservation and conservation purposes. This study combines acoustic (Sonar) and laser-based (Lidar) remote sensing technologies in an aquatic environment with two machine and deep learning approaches to illustrate the techniques to identify submerged objects. Firstly, the relative accuracy of the underwater imaging system, the BlueView BV5000 Mechanical Scanning Sonar, is evaluated at close range. Secondly, the supervised CANUPO and RandLA-Net classification approaches are used to classify submerged sonar point clouds. Common objects of interest, namely tyres and chairs, were selected for classification. Relative accuracy measurement results showed a centimetre-level root mean square error (RMSE) value, with good accuracies recorded when the scanner is positioned close to objects. The best results were achieved when the target objects were placed at a minimum distance of 2 m from the acoustic scanner. Subsequently, the results of point cloud classification were satisfactory for both approaches. An overall accuracy of 79.81% and an F_1 score of 79.80% were achieved using the CANUPO classification approach. On the other hand, an 80.72% overall accuracy and an 80.63% F_1 score were obtained using a RandLA-Net approach. These analyses provide a reasonable framework for the parameters that can be used when applying these techniques in natural aquatic environments.

Keywords: *3D Sonar point clouds, Accuracy assessment, Point cloud classification, CANUPO, RandLA-Net, Underwater Cultural Heritage*

1. Introduction

The classification of three-dimensional (3D) point cloud data using machine and deep learning (ML and DL) approaches has been a centre of interest in terrestrial environments. A large collection of points characterises point clouds, each defined by XYZ coordinates in a three-dimensional (3D) space. Additional information, such as reflectivity and colour values, may be contained in a point cloud depending on the sensor's mode of measurement (Van Genechten, 2008). Owing to

advancements in point cloud acquisition technologies, light detection and ranging (Lidar) and structure-from-motion (SfM) photogrammetry, these approaches have become solutions in acquiring point cloud data (Hu *et al*., 2012; Westoby *et al*., 2012). As a result, numerous studies have classified terrestrial point clouds for mapping (Alexander *et al*., 2011; Peeroo *et al*., 2017; Ntuli and Forbes, 2023), disaster management (Richter *et al*., 2013) and preserving cultural heritage sites (Grilli *et al*., 2019; Pepe *et al*., 2022). While the aforementioned studies have focused on classifying point clouds above the earth's surface, research is expanding to explore underwater environments. In the point cloud context, classification is assigning predefined category labels to a group of points in a point cloud (Weinmann *et al*., 2015). A limited number of studies have explored the classification of underwater objects for various applications. Examples include the recovery of lost objects (Tsai *et al*. (2021), terrain-based navigation, target object detection (Sung *et al*. (2020), damage detection (Hake *et al*. (2023), and navigation (Palomer *et al*., 2015). Most studies rely on the geometry of point clouds for classification since acquiring other attributes in the underwater environment may be challenging. Among the intriguing prospective uses of underwater point cloud classification is the identification and preservation of the underwater cultural heritage. In Liarokapis *et al*. (2017), optical data were employed to create models and maps of submerged archaeological assets. This methodology facilitated the derivation of precise archaeological information on the shipwreck and amphorae which was rendered available in virtual reality. In addition to recording and conserving historical sites, underwater archaeology provides insights into past civilisations (Bowens, 2011). South Africa is rich in underwater archaeology, mainly, in the form of shipwrecks settled beneath our oceans. However, very few sites have been surveyed and documented (DSAC, 2009). There is also a necessity to classify and document archaeological objects submerged underwater.

Menna *et al*. (2018) have provided a comprehensive overview of the state-of-the-art techniques and applications of the 3D recording and mapping of the underwater cultural heritage. Various technologies and methods, including laser scanning and acoustic imaging, for surveying underwater archaeological sites have been discussed. Among these technologies is the Blueview multi-beam echosounder, developed by Teledyne Marine (2018), which is capable of acquiring 2D sonar images. With the advancing developments, a 3D Mechanical Scanning Sonar (MSS) BV5000 has been introduced (Teledyne Marine, 2020). The BV5000 is an MSS that produces high-resolution imagery and point clouds of underwater objects and scenes. It can be mounted on a remotely operated vehicle (ROV) or tripod and produce both sector and spherical scan data. 3D scanning capabilities are achieved even under low or zero visibility conditions. Acoustic signals are preferred underwater as they can effectively propagate and reflect, making them less vulnerable to environmental conditions (Forrest, 1994). According to UNESCO (2015), non-destructive survey methods are preferred over object recovery to protect the underwater cultural heritage. Moreover, non-invasive methods should be employed when excavation is required to conserve the artifacts. Hence, the BV5000 MSS is suitable for such applications. Other studies have assessed the ability of the BV5000 MSS to measure objects accurately. For example, Moisan *et al*. (2018) conducted a study that compared the point clouds produced by the BV5000 MSS with those produced by a Terrestrial Laser Scanner (TLS) in a

harbour lock. When empty, a lock was scanned using a TLS and then filled to perform the survey using the BV5000 MSS. The results revealed that detecting details smaller than 4cm was more difficult using the BV5000 MSS. Another study by Lesnikowski and Rush (2012) from the Blueview Corporation showed a similar comparison when the accuracy was evaluated in a dry dock. Initially, a TLS survey of the vacant dry dock was conducted; sonar acquisitions were then performed once the dock had been filled. Differences of 3-5cm were obtained for translational measurements. Additionally, the results revealed that better outcomes are achieved when the Blueview MSS is positioned in close proximity to the target object. Similarly, we evaluated the relative accuracy of the BV5000 MSS in close range by comparing measurements at different distances from point clouds of the same objects gathered on land and underwater.

Extensive work has been invested in classifying point cloud data using deep learning methods. PointNet++ and PointCNN networks have shown remarkable competence, outperforming state-of-the-art techniques in this field (Qi et al., 2017; Li et al., 2018). Several research works have classified underwater-derived point clouds for various applications. Sung *et al*. (2020) employed a neural network to train and classify a point cloud created from sonar images captured through an autonomous underwater vehicle (AUV) by reconstructing the 3D shape of the object. The method proposed is suitable for navigating across the terrain and identifying objects. In Hake *et al*. (2023), underwater point clouds were converted to depth images and classified using a neural network to automate damage detection in port structures. Moreover, Tsai *et al*. (2021) conducted a study wherein they implemented neural networks (PointNet and PointConv) described in Qi *et al.* (2017) and Wu *et al*. (2019) to classify a 3D point cloud underwater to identify discarded tyres. The results demonstrated promising potential for detecting lost objects in underwater environments. Himri *et al*. (2019) conducted a comparative analysis of state-of-the-art techniques for object recognition leveraging the open-source Point Cloud Library (PCL) (Rusu and Cousins, 2011). This study indicated the challenges in underwater environments, including noise and fluctuations in resolution. Furthermore, the study underscored the critical significance of aligning the resolution of both the database and test scan datasets to optimise descriptor performance. While previously mentioned studies have significantly contributed to underwater point cloud classification, in this research, we explore the multi-class classification of sonar point clouds using CANUPO and RandLA-Net approaches, evaluate the performance of these approaches and compare the outcomes.

2. Study Area

On account of several compelling factors, the swimming pool at the University of KwaZulu-Natal (UKZN) Westville Campus was identified as the study area for the underwater point cloud classification research. The controlled environment of the swimming pool provides the best conditions for conducting experiments and minimising external factors that could influence study outcomes. Moreover, the swimming pool is accessible and safe, and allows for an adequate time frame for deploying underwater imaging technologies for effective data collection. Figure 1 shows

the UKZN's Westville campus swimming pool, from which training and testing datasets were collected.

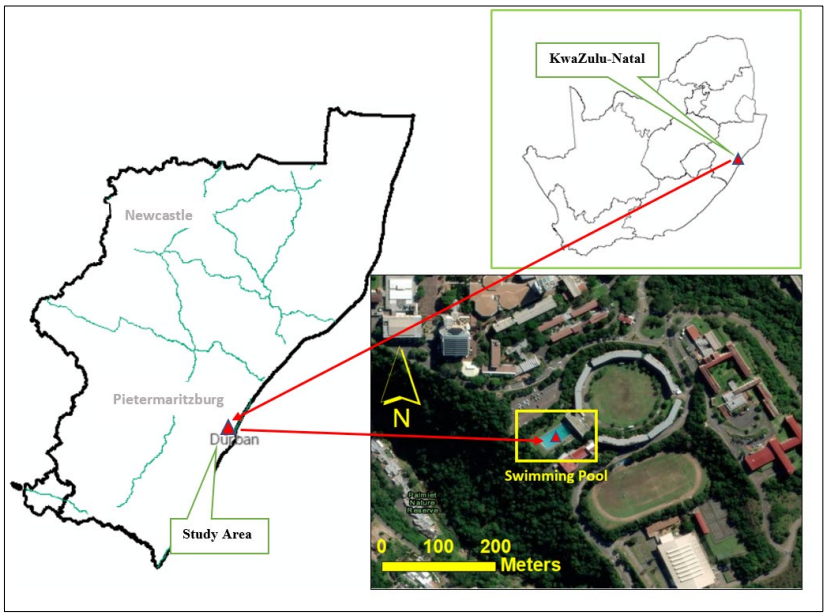

Figure 1: Presentation of the study area in KwaZulu-Natal Province, Aerial View (WGS84 coordinates of the red triangle: 29°49'12.35" S, 30°56'43.19" E).

3. Materials and Methods

3.1. Description of Instruments and Objects

Two types of scanners, namely the acoustic Sonar multibeam scanner and the terrestrial laser scanner, were used for accuracy assessments and point cloud acquisitions. The Blueview® (www.blueview.com) BV5000 mechanical scanning sonar (MSS), pan and tilt system, tripod, and associated accessories were used (see Figure 2(a)). The Sound Navigation and Ranging Sonar measures distances by using the propagation of sound waves which makes it resistant to changes in lighting. This makes it particularly suitable for applications in challenging underwater environments. The BV5000 Mechanical Scanning Sonar (MSS) is a specialised underwater imaging system that operates at a frequency of 1.35 MHz, making it possible to capture high-resolution 3D point clouds of underwater structures and environments. It operates in a similar manner to that of a Terrestrial Laser Scanner (TLS), but instead of using light waves for measurement purposes, it employs sound waves. Moreover, the system emits fan-shaped sonar signals and receives echoes, creating a comprehensive 3D point cloud while panning around an object or an underwater environment. The terrestrial laser scanner, a Faro 3D Focus 120 TLS (see Figure 2(b)) was used to collect 3D point cloud data on land. This instrument is phase shift-based and may collect data within a 0.6 m – 120 m range. It emits a laser beam by using a rotating mirror, which is directed toward the scanning area and covers a vertical and horizontal range of 305º and 360º, respectively. As the laser beam encounters objects, it reflects back to the scanner, and the distances to these objects can then be

determined. A comparison of the characteristics of the BV5000 MSS and the Faro Focus 3D 120 TLS is presented in Table 1 below.

Table 1: The characteristics of the Blueview BV5000 and the Faro FOCUS 3D 120 TLS as per the specifications of the manufacturer

	Blueview BV5000 Sonar	**Faro FOCUS 3D 120 TLS**
Beam Width	1º x 1º (mm)	3 mm
Ranging error	15 mm	±2 mm (10-25 m)
Maximum Range	30 m	120 m
Horizontal resolution	~ 0.09º (16 mm@10 m)	0.009º
Vertical resolution	0.18º (30 mm@10 m)	0.009º
Field-of-view (vertical/horizontal)	45º /360º (320º /360º)	305º /360º

To facilitate the assessment of the relative accuracy of the BV5000 MSS the relevant metal and wooden cubes and cuboids employed in the research study. These objects were manufactured according to strictly prescribed dimensions so as to minimise the measuring errors (see (Figure 2 (c) and (d)). The cubes measured 40cm on each side, while the vertices of the rectangular cuboids measured 40cm, 50cm and 70cm, respectively. Furthermore, plastic chairs and rubber tyres (see Figure 3(e) and (f)), were used for the point cloud classifications. These objects have good reflective properties and are commonly found in the real world. Several software programmes were used in this investigation. ArcMap 10.6 was used for map creation; ProScan 3.13 and ProViewer were used to configure the scanner and to collect data; Blueviewer software was used to view 3D point cloud data on site; and CloudCompare software (Girardeau-Montaut, 2023) was used for point cloud processing − the last-mentioned including noise filtering and scan registration; and ArcGIS Pro 3.1 and CloudCompare 2.13 beta were used to perform point cloud classifications.

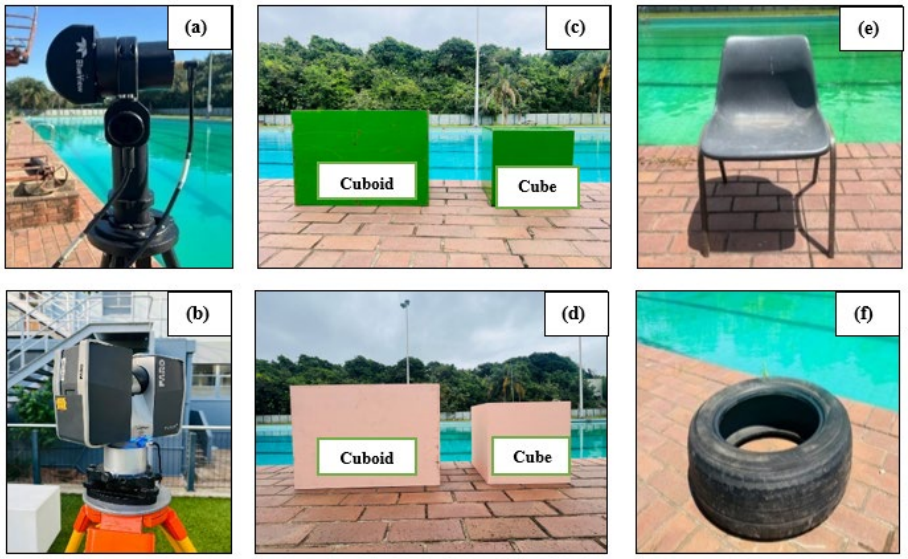

Figure 2: (a) Blueview BV5000 MSS; (b) Faro FOCUS 3D 120 TLS; (c) metallic cuboid and cube; (d) wooden cuboid and cube (e) plastic chair; and (f) rubber tyre.

3.2. Methods

The methods used in the study are explained in more detail in this section. The stages are as follows: point cloud data acquisition, where the point clouds were acquired using both Lidar and acoustic scanners; point cloud cleaning and registration, where the noise was removed and individual scans were combined to create a full 3D view of the scene; scanner accuracy assessment, where the accuracy of the acoustic scanner was evaluated against the measurements from the Lidar scanner; point cloud classification, where CANUPO and RandLA-Net classification approaches were applied to classify underwater point clouds; and classification accuracy assessment, where the performance of the classification models was evaluated. Figure 3 illustrates the comprehensive procedures adopted throughout this study.

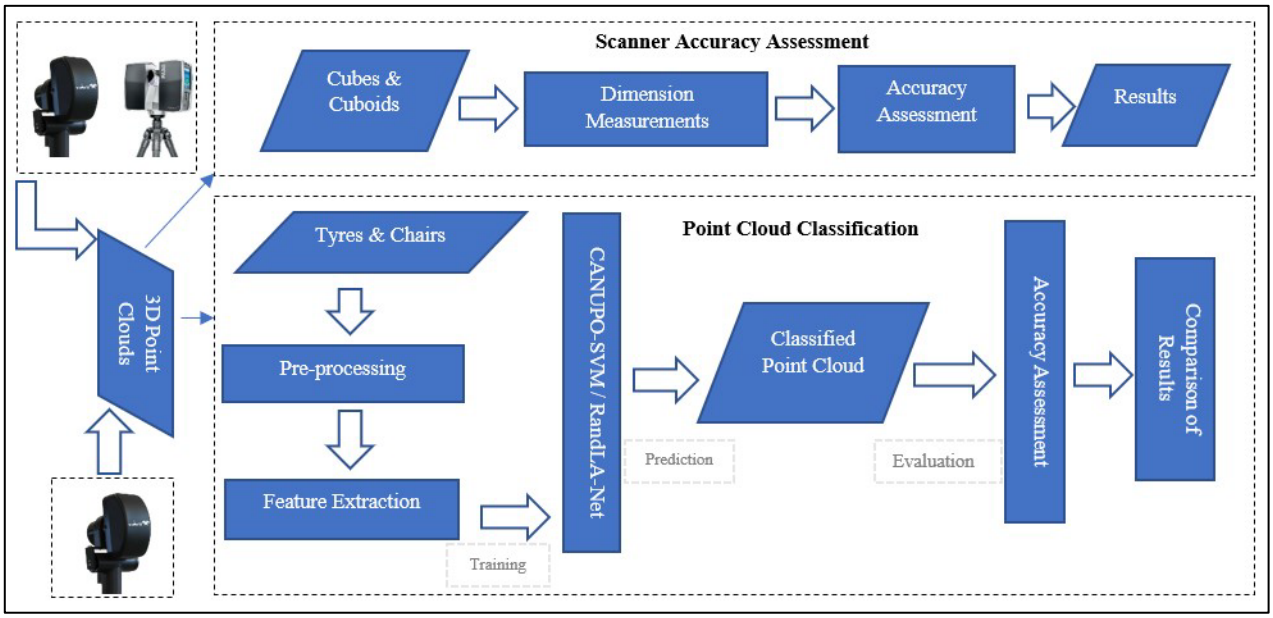

Figure 3: Flow chart showing the overall methodology workflow: accuracy assessment and point cloud classification.

3.2.1. Point Cloud Data Acquisition

The first dataset of this study was used to assess the relative accuracy of the Blueview BV5000 MSS at close range. A cuboid and a cube were scanned using the Faro Focus (360) TLS at distances of 2m, 4m, and 6m, respectively. Similarly, while submerged in the swimming pool, the cuboid and cube were scanned at exact distances by the BV5000 MSS. Point cloud data were collected using both laser and acoustic scanning processes. The linear surfaces of the cubes and cuboids reflected the laser and acoustic signals to the effect that their respective outlines were clearly visible on the point clouds. Following the data collection process, a manual cleaning process was performed on the point clouds to eliminate unwanted noise. Furthermore, corresponding points and dimensions were identified for distance measurements and comparisons.

The second dataset applied in of this study was used for the classification of 3D Sonar point clouds in the sub-aquatic environment. 3D point cloud data were collected using a Blueview BV5000 MSS mounted on a tripod and connected to ProScan software. Figure 4(a) shows the experimental setup

before the researcher collected the 3D point cloud data. Initially, for classification purposes, only the chairs and tyres were submerged in the swimming pool. This was followed by the immersion of the BV5000 MSS in the pool until the tripod was firmly settled on the floor. Image calibration was then performed to ensure accurate measurements. Each scan adhered to identical scanning parameters, namely one detection per beam and the employment of spherical scans covering a 360° horizontal range, along with vertical angles of 0° and -15°. In some instances, and depending on the position of the objects, a vertical angle of -45° was used. Multiple scanner positions were used to obtain a complete 3D scene image. However, although classification models should function well with only a single scan, five scans were collected for each dataset. Figure 4(b) shows the 3D point cloud produced by the BV5000, with the scanner positioned at the centre of the objects. The raw 3D point cloud data contained isolated points resulting from reflections of the pool walls and surface. Point cloud data cleaning was performed manually to remove noise brought on by backscattering.

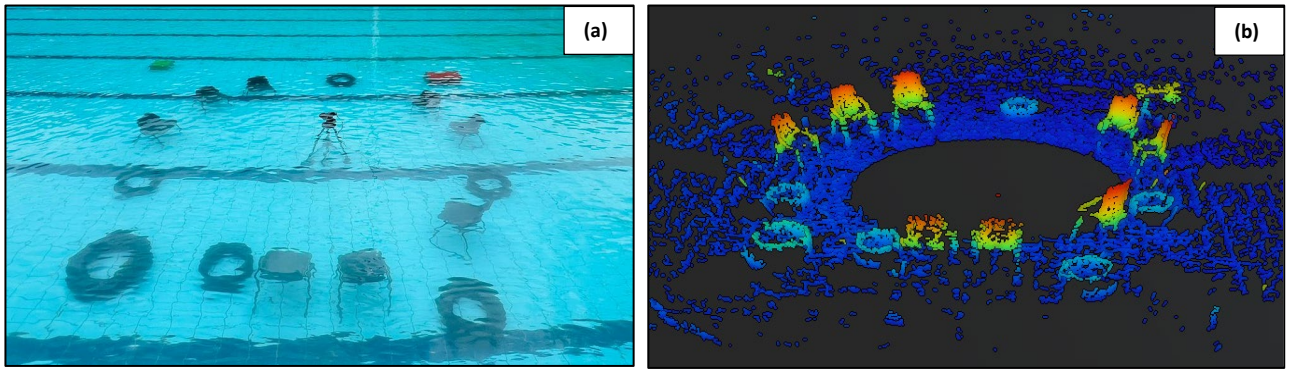

Figure 4: (a) Objects submerged in water; (b) Corresponding 3D point cloud data collected using the BV5000 MSS

3.2.2. Point Cloud Cleaning and Registration

CloudCompare software was used to process the point cloud data collected through the Lidar and acoustic scanners. For the underwater environment, the noise was filtered to remove the outliers and the swimming pool floor reflection data from the point cloud data. The interactive segmentation tool in CloudCompare was used to remove outliers in the point cloud data. Data points identified as noise were manually removed from the point cloud, while the noise filter automatically removed the outliers. Furthermore, the CSF algorithm, devised by Zhang *et al*. (2016), was also used to remove the swimming pool floor reflection data. The data collected had sufficient overlap to allow point cloud registration in that a cloud-based method could be applied. Scans collected from different positions were aligned to obtain a complete 3D scene through registration. This approach was time-consuming; however, good accuracies were obtained. In addition to CloudCompare, it should be noted that alternative software applications, such as AVEVA Point Cloud Manager (AVEVA, 2020) and Leica Cyclone (Leica Geosystems, 2017), are available for the purposes of point cloud cleaning and registration.

3.2.3. Scanner Accuracy Assessment

To evaluate the relative accuracy of the BV5000 MSS, measurements were taken on selected objects of interest, namely a cuboid and a cube made up of metallic and wooden materials. Point cloud data were acquired in the outdoor environment using a Faro Focus 3D 120 TLS, starting from a distance of 2 m from the objects. Data were then collected at 4 m and 6 m distances, respectively. The same objects were submerged in water to acquire underwater point cloud data with the aid of the BV5000 MSS and by applying the same experimental procedure. Inherent challenges, such as drifting and floating objects associated with the underwater environment, emerged. However, weights were used to ensure that the objects remained stationary while underwater. Numerous measurements were taken using each of the respective instruments. The corresponding points and dimensions were identified in the point cloud from the reflections of the surfaces of the cuboids and cubes, thereby allowing for accurate measurements of the scanner residuals to be made. TLS point cloud data were used as reference models from which expected values could be measured. Lastly, the consistency and the concurrence between the measurements acquired by the two scanners were analysed by determining the differences in the measured values. Furthermore, the Root Mean Square Error (RMSE) (equation 1) of the linear measurements was used to determine the relative accuracy for each distance (Willmott, 1982).

$$RMSE = \sqrt{\sum_{i=1}^{n} \frac{(\hat{x}_i - x_i)^2}{n}} \quad [1]$$

Where:

$RMSE$ - Root Mean Square Error

$\hat{x}_i$ - Predicted value

x_i - Observed value

n - Number of observations

3.2.4. Point Cloud Classification

The main objective of this study lay in exploring Machine and Deep Learning algorithms for classification tasks in the underwater environment. Machine learning aims to equip computers with the ability to identify patterns through data-driven learning, thereby enhancing their performance without the need for explicit programming. Moreover, computational algorithms and techniques are developed to automate the recognition process and save time (Langley and Simon, 1995). On the other hand, Deep Learning is a progression of machine learning, where algorithms are arranged in layers to create an artificial neural network capable of autonomous learning and intelligent decision-making (Wang and Raj, 2017). For this study, two distinct classification algorithms were selected, namely, CANUPO and RandLA-Net. These algorithms were selected with due consideration given to their widespread adoption and proven efficacy across diverse classification tasks, even though, to

the best of our knowledge, they have not been previously employed for underwater point cloud classification.

The sonar point clouds do not contain colour information; thus, the geometry of points is considered for feature extraction. The datasets remained consistent throughout, with identical datasets used for training across both classification algorithms, with 70% allocated for training and 30% for validation purposes. Additionally, an independent testing dataset with similar characteristics was used. The steps involved in the classification process are outlined in Figure 3. The training dataset comprised point clouds containing chairs and tyres carefully annotated for supervised classification. Another dataset was reserved for validation to check whether the models could generalise on a labelled dataset and to assess the accuracy of the models. Lastly, the testing dataset comprised unseen point clouds for checking whether the models could in fact generalise in new environments.

3.2.5. CANUPO Classification

Cortes and Vapnik (1995) developed the Support Vector Machine (SVM) learning algorithm, which is currently used on a large scale in classification and regression tasks. This algorithm employs a predetermined non-linear mapping function to transform the input vectors into a higher-dimensional feature space. It is then possible to establish a linear decision surface with special attributes that guarantee the network's exceptional ability to generalise effectively within this feature space (Cortes and Vapnik, 1995). The SVM aims to identify a hyperplane that not only separates the categories but also maximises the margin between the closest data points of different classes, known as support vectors (Boser *et al*., 1992). As a result, the SVM is robust in over-fitting and generalises well in respect of unseen data. It has demonstrated its strong performance, even when trained on small datasets (Cortes and Vapnik, 1995; Foody and Mathur, 2004). In this study, we employed the CAractérisation de NUages de POints (CANUPO) - SVM classification algorithm, developed by Brodu and Lague (2012). This algorithm relies on the local dimensionality attributes of points within a point cloud. Depending on the chosen scale and location, the points are categorized as 1D, 2D, or 3D. Combining information on various scales results in distinctive signatures, enabling the recognition of different object classes within the scene. These signatures are automatically generated through training points, facilitating the optimisation of class separability. Moreover, descriptors are created by integrating information from different scales, thereby aiding in identifying other object classes present on the scene. The classification process involves two steps: firstly, the data are projected onto a plane to achieve maximum separability, and secondly, classes within that plane are divided by a boundary. In this study, the CANUPO algorithm was executed using CloudCompare software, with the classification parameters set to dimensionality. Since CANUPO is based on a binary classification, the classifier was trained to identify only tyres and chairs using 10 000 maximum core points and scales between 0.4m and 1.00m.

3.2.6. RandLA-Net Classification

RandLA-Net is a Deep-learning (DL) neural architecture designed for the semantic segregation of large-scale 3D point clouds. This model employs a random sampling method to reduce the point

density in extensive datasets (Hu *et al*., 2020). The architecture of the RandLA-Net network comprises input, encoding, decoding, and output layers. The input layer processes the point cloud data containing 3D coordinates, with the possibility, in some instances, of including colour information. This input layer then connects to the encoder-decoder layer with skip connections. Each encoding layer employs the random downsampling of the point cloud, thereby significantly reducing its point density. The network incorporates a local feature aggregation module for each point to preserve the essential features of the downsampled point cloud. Each decoding layer then performs an upsampling of the point feature set by using nearest-neighbour interpolation. The network combines these upsampled maps with feature maps generated by the encoding layers through skip connections and applies a shared MLP to the concatenated feature maps. Finally, the output layer provides predictions for all points in the point cloud (Hu *et al*., 2020).

The RandLA-Net algorithm was executed in ArcGIS Pro version 3.1 Software and the classification procedure was divided into four steps (ESRI, 2023). The first step involves data preparation whereby the point cloud data for training and validation are converted into an HDF5 format to create training data (The HDF Group, 2023). These data are then used in the second step to train a DL model for the point cloud classification task. The third step assesses the quality of the point cloud classification model using an already annotated point cloud. Lastly, the developed model is applied to classify the unseen point cloud.

Both training and validation data were extracted from the same point cloud by splitting the data into two datasets. An additional point cloud, comprising the same objects of interest, was used to test the RandLA-Net model. The model was trained using 15 epochs with 100% of the iterations per epoch and a batch size of two. Three classes were pre-assigned, namely, tyre, chair, and unassigned. The unassigned class was excluded during the preparation and training steps. As a result, only the tyres and chairs were considered.

3.2.7. Classification Accuracy Assessment

The performance of classification approaches used in this study was evaluated using confusion matrices. The rows and columns in this matrix represent the true and predicted class labels, respectively. For a more in-depth analysis, secondary indicators derived from statistical results were used. Accuracy, Precision, Recall, and the F1 score gave insightful information about the performance and accuracy of the classification techniques. The precision of a classifier is the percentage of true positives, or positive predictions, that are in fact correct. Contrarily, recall (sensitivity) represents the proportion of true positives correctly identified as positive by the classifier. Lastly, the F1 score indicates the overall accuracy of the model. From the confusion matrix, the following metrics were obtained:

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \quad [2]$$

$$Precision = \frac{TP}{TP + FP} \quad [3]$$

$$Recall = \frac{Tp}{TP + FN} \quad [4]$$

$$F1\ score = \frac{Recall * Precision}{Recall + Precision} \quad [5]$$

Where:

TP - true positive

TN - true negative

FP - false positive

FN - false negative

4. Results and Discussion

4.1. Results of Scanner Accuracy Assessment

Measurements obtained from the Lidar point clouds were used as references to assess the accuracy of the BV5000 MSS. Multiple measurements were taken to observe good precision, and like the method employed by Coetzee and Singh (2021), the average values were calculated. Figure 5 shows the Root Mean Square Error (RMSE) values calculated for each distance, indicative of the mean of discrepancies between the estimated and observed values. A trend of increasing RMSE values was noticed as the distance increased. This is consistent with the difficulties that are expected as measurement ranges are extended. For example, it was found that there was a loss of energy from the acoustic and Lidar scanners, respectively, as the distance from the scanners increased. This is known as signal attenuation. There was also absorption and scattering of the signal from air and water particles. Interestingly, the results show the scattering properties of wooden and metal objects. The reflective surface in the point cloud data generated from the wooden materials had a higher density of points when compared to the reflections from the metallic objects. As a result, the task of measuring the dimensions of point clouds was more challenging in the case of the metallic objects. At distances of 2m and 4m from the scanner, we observed that metal and wooden objects present with negligible RMSE differences of ±4cm. At 6m from the scanner, it was observed that wooden objects have lower RMSE values than metal objects. This could be due to the scattering of the signal from the surfaces of the metal objects as opposed to that from the wooden objects, that would, in the latter case, provide better reflections. This also shows that the optimal measuring distance from the scanner is 4m for scanning objects with dimensions of 40cm and larger, thereby resulting in errors of ±4cm.

Figure 5 also shows the different scattering properties of the cubes and cuboids. No difference can be seen at 2m from the scanner, while the RMSE differences become more prominent at distances of 4m and 6m from the scanner. In general, this could mean that smaller RMSE differences are observed when assessing reflections from the cube as opposed to those from the cuboid. These differences are primarily due to the smaller dimensions of a cube as opposed to those of a cuboid.

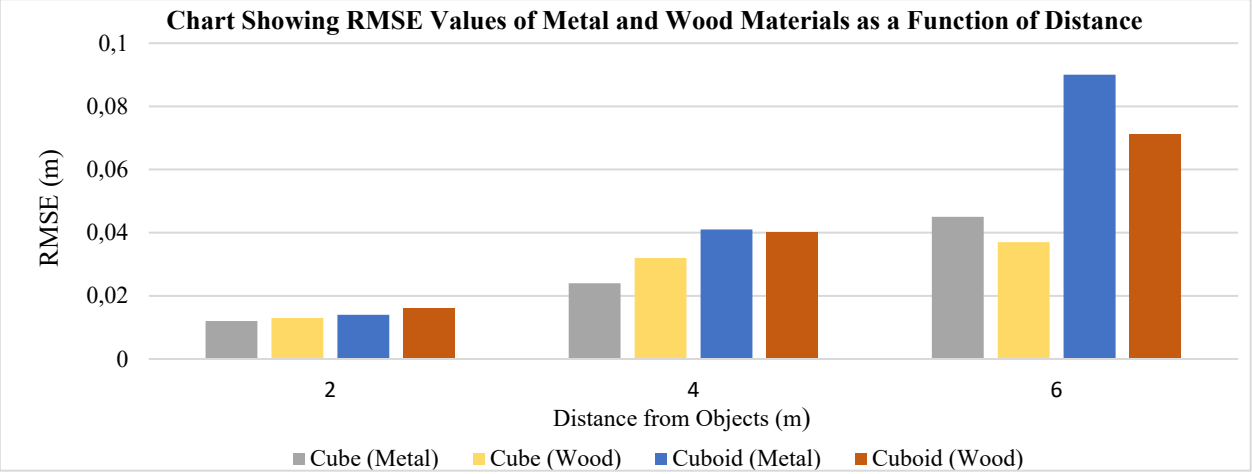

Figure 5: Chart showing RMSE values as a function of distance

It was observed that the primary challenge with increasing distance was the density of points, resulting in difficulties when taking measurements. Moreover, in some instances, objects were not fully imaged, resulting in point clouds with discontinuities. These results reveal that placing the SONAR scanner closer to the target objects produces scans of high-density points; hence, a better resolution. As a result, the ideal distance for applications in close-range conditions is determined by the geometric accuracy assessment. Therefore, the BV5000 MSS was positioned approximately 2m from the target objects during point cloud acquisition for classification tasks. The outcomes of this research are comparable to what has been documented in other studies, such as in the research studies of Coetzee and Singh (2021) and Moisan *et al*. (2018).

4.2. Results of Classification and Model Comparisons

In this study, the CANUPO and RandLA-Net approaches were used to classify objects of interest, namely tyres and chairs, from the submerged point cloud. The CANUPO approach is based on the dimensionality of object points and is capable of combining different scales to create descriptors. Considering the size of the target objects, the selected scales ranged between 0.40m and 1.00m. During the classifier training, high separability and recognition values were achieved. Moreover, to determine whether the created model could make generalisations on unseen data, an independent point cloud was used for model testing. The swimming pool floor was removed and the classifier was applied to the remaining points.

Figure 6 shows the experimental setup and the datasets that were used to create the classifier. Furthermore, the CANUPO classification results are also presented in Figure 6(d). A visual inspection of the point cloud classification findings shows that the tyre category achieved a high recognition rate. However, this category was affected mainly by commission errors. This could be due to the geometry of the chair legs and the hollow shape resembling the tyres used in the training dataset. Some points belonging to the chair category were classified as tyres. On the other hand, since the binary classification was applied, the chair category was influenced by errors of omission. A confusion matrix was applied to quantitatively evaluate the performance of the CANUPO model (see Table 2).

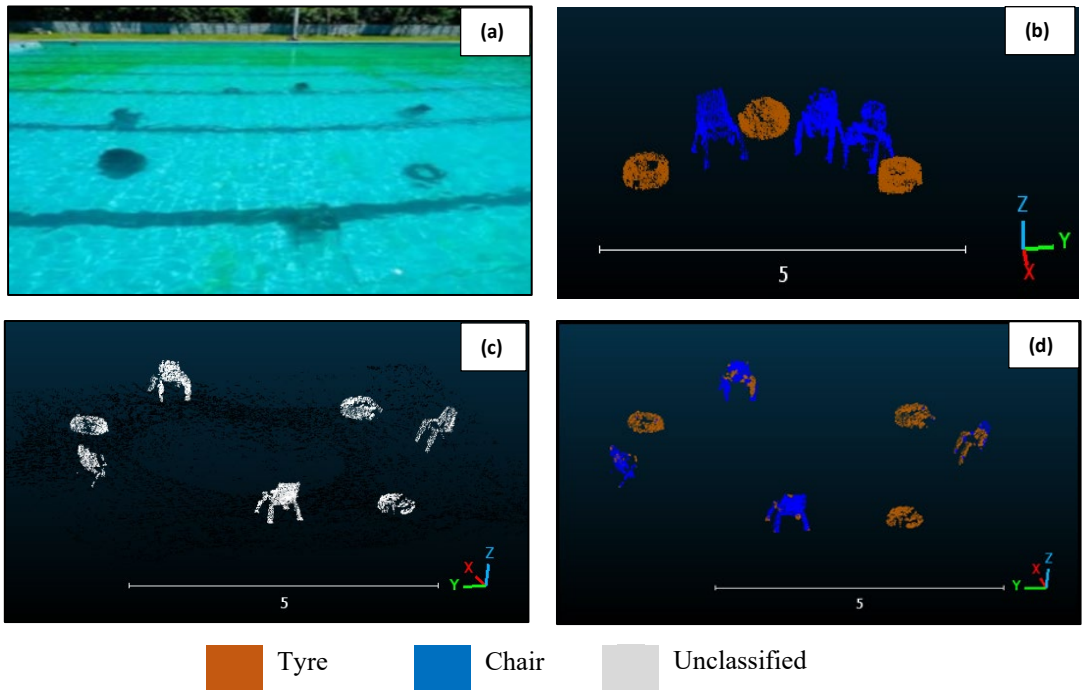

Figure 6: Classification results of the CANUPO Model: (a) An image of the test dataset (b) Training dataset (c) Test dataset (d) Classified test dataset.

Table 2: Classification evaluation results of the CANUPO Model

Class/Model	**Training Set**			**Test Set**		
	Precision	Recall	F1 Score	Precision	Recall	F1 Score
Tyre	0.930	0.967	0.948	0.987	0.680	0.790
Chair	0.947	0.927	0.946	0.680	0.989	0.806
Overall accuracy	94.72%			79.81%		

Table 2 shows the quantitative evaluation results determined using a confusion matrix. During the training phase, an overall accuracy of 94.72% was achieved, thereby demonstrating how well CANUPO can differentiate between chairs and tyres in point cloud data. Moreover, upon applying the model to the unseen dataset, an overall classification accuracy of 79.81% was achieved. With recall at 68% and precision at 98.7% in the tyre category, the F1 score was 79%. Conversely, the chair category received an F1 score of 80.6%, with a precision of 68% and a recall of 98.9%.

Figure 7 presents the classification results obtained using the RandLA-Net model. The same training data used in CANUPO was used. In (a), the reference evaluation set is shown, and in (b), the classification results of the validation data are displayed. The model was trained at 15 epochs, and the points that were part of the reflection from the pool floor were excluded during both the training and validation phases. As in the testing of the CANUPO model, the RandLA-Net model was tested on unseen data to evaluate the accuracy of the classification. Upon visual inspection, Figure 7(b) shows the classification results on the validation set. A few misclassifications in the chair category

were significant. On the other hand, the tyres were associated with a high degree of recognition. On the test dataset, the tyre category was associated with commission errors. Furthermore, a quantitative assessment of classification accuracy was conducted. In Table 3, an overall accuracy of 94.08% was achieved on the validation dataset.

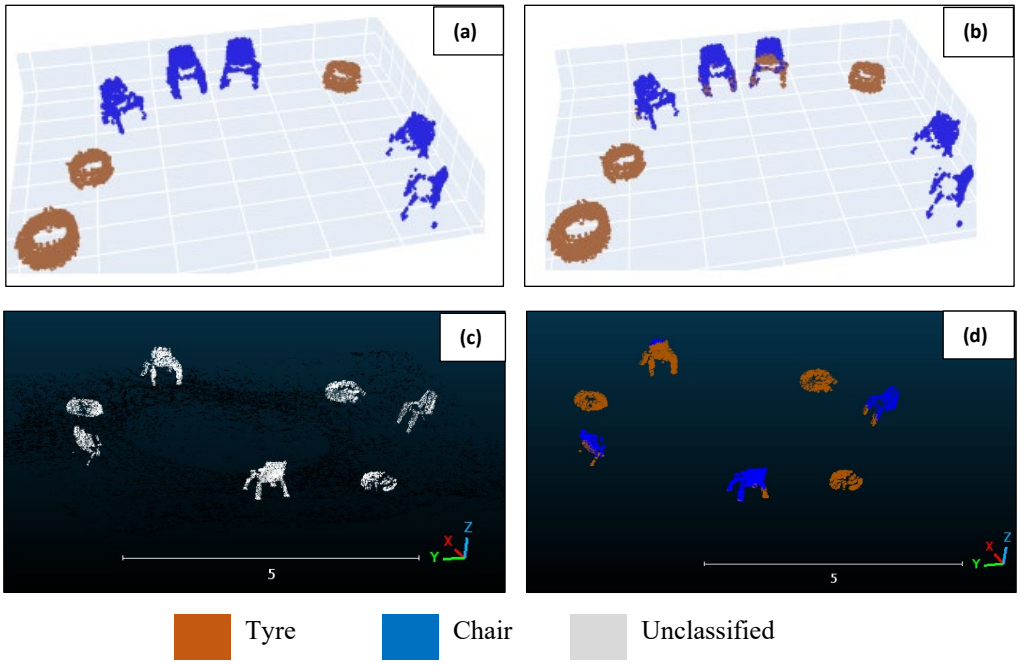

Figure 7: Classification results of the RandLA-Net Model: (a) Validation dataset, (b) Classified validation dataset, (c) Test dataset and (d) Classified test dataset.

Table 3 Classification evaluation results of the RandLA-Net Model

Class/Model	**Validation Set**			**Test Set**		
	Precision	Recall	F1 Score	Precision	Recall	F1 Score
Tyre	0.904	1.00	0.950	0.657	0.999	0.793
Chair	1.00	0.866	0.928	0.999	0.695	0.819
Overall accuracy	94.08%			80.72%		

An overall accuracy of 80.72% was achieved by testing the RandLA-Net model on an unseen dataset. Regarding tyres, 65.7% precision, 99% recall, and a 79.3% F1 Score were recorded. In contrast, the chair category received an F1 Score of 81.9%, 69.5% recall, and 99% precision. Table 4 provides a summary of the classification outcomes for each method. Comparable results were obtained during the validation phase, and 94.72% and 94.08% were achieved for the CANUPO and RandLA-Net models, respectively. These results indicate the ability of both models to excel in respect of the training data. Nevertheless, testing resulted in a significant accuracy decline to 79.81% for the CANUPO and 80.72% for the RandLA-Net model, indicating possible challenges in making generalisations in respect of unseen data. These results show accuracy comparable to underwater point cloud classification research, including that conducted by Tsai *et al*. (2021) and Kogut and

Slowik (2021). However, Tsai *et al*. (2021) achieved higher accuracies using the PointNet and PointConv network architectures. These studies focused only on classifying one category at a time. In this study, two classification categories were considered.

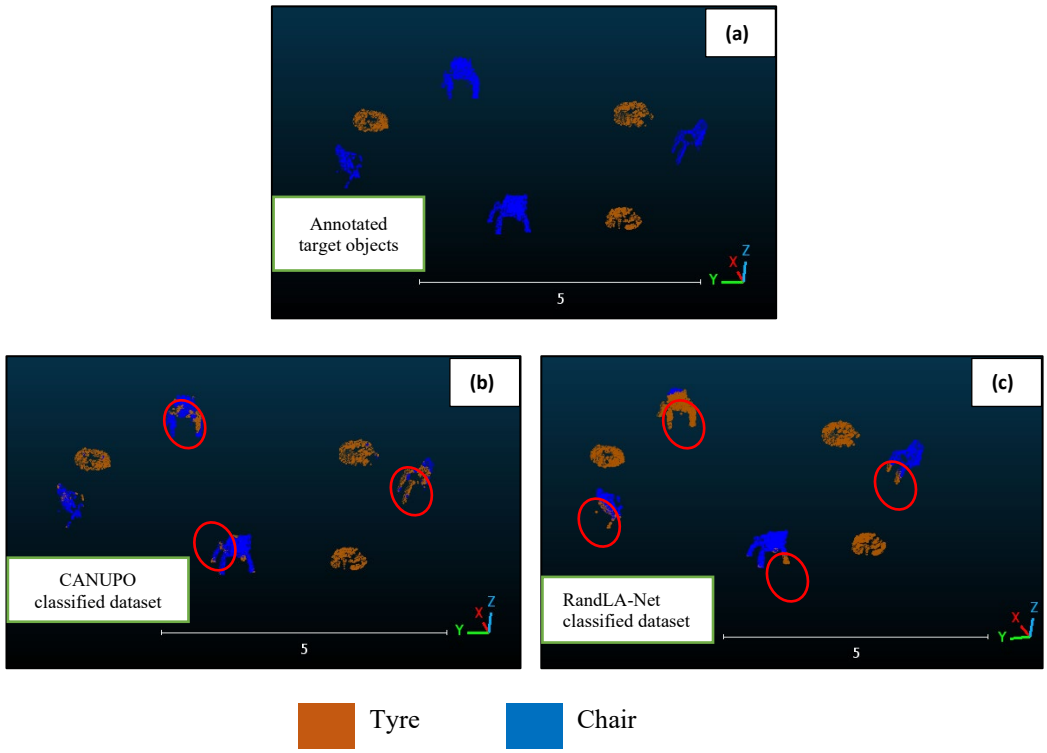

Figure 7: (a) Ground truth test dataset, (b) Classification results of the CANUPO Model and (c) Classification results of the RandLA-Net Model

Table 4: A summary of classification evaluation results of CANUPO and RandLA-Net

	CANUPO		**RandLA-Net**	
	Validation Data	Test Da	Validation Data	Test Data
Precision	93.85%	83.35%	95.20%	82.81%
Recall	94.70%	83.45%	93.30%	84.73%
F1 Score	94.70%	79.80%	93.89%	80.63%
Accuracy	94.72%	79.81%	94.08%	80.72%

The CANUPO model took about two minutes to train, while the RandLA-Net model took an hour and 10 minutes on the CPU. Both models achieved satisfactory classification results, although significant misclassifications (circled in red in Figure 7) were also noticed when the models were applied to unseen data. The possible reasons behind these results include the limited datasets used for training. To train classification models that perform well with unseen data, it is essential to use a diverse training dataset. Other studies have used data augmentation to overcome this problem (Tsai *et al*., 2021; Long *et al*., 2023). Apart from limited datasets, underwater environments are dynamic, and the quality of the point clouds acquired is subject to noise owing to attenuation of the signal.

Moreover, particularly in this research, in the case of the chair category, point clouds were associated with discontinuities. Certain legs and chair seats were not fully reconstructed. To keep variables constant, an optimal distance of 2m from the scanner was used to eliminate the attenuating effect of the signal. It was also noted that the two objects were made of different materials: the plastic chair, with steel legs, and the rubber tyre. The difference in the material properties of the objects caused negligible differences in the effectiveness of the performance of both classification approaches. The classification errors in this research are possibly due to the geometric effects, where, as a result of the shadow effect, the reflections from the chair resemble those of the tyre. The developed models performed well in classifying objects, even when used on point clouds obtained from a single scan. This highlights the ability of the model to accurately recognise and classify objects in the context of a single scan.

These findings point to the potential for improvement, especially considering the limitations of small datasets. Opportunities for improvement include a more comprehensive range of training datasets. Furthermore, beyond the CANUPO and RandLA-Net methodologies, other pertinent software tools contribute significantly to point cloud processing and classification. Notable examples include Pointly (Pointly, 2024), Lidar360 (Lidar360, 2013), The Cross Product (The Cross Product, 2020) and Pix4D Mapper (Pix4D SA, 2024).

5. Conclusion

The first aim of this research was to evaluate the relative accuracy of the BV5000 MSS, while the second aim was to classify submerged sonar point clouds. This study has reported on the relative accuracy of the BV5000 Mechanical Scanning Sonar on metal and wooden materials for close-range application in a controlled underwater environment. At varying distances, objects of different sizes and materials were scanned on land using a TLS. Moreover, a similar setup was replicated in the underwater environment. The results indicate that positioning the BV5000 MSS closer to the target objects provides measurements that are comparable with those obtained using a TLS. Furthermore, wooden objects had higher point cloud densities and resolution than metal objects. Owing to the possible scattering effects, the wooden objects had lower RMSE values at greater distances than the metallic objects. These findings provide valuable insights into determining the optimal scanner positions for the classification task.

The second aim of this research extended to point cloud classification using CANUPO and RandLA-Net approaches. Underwater point cloud classification has gained popularity in various industry applications. As regards this contribution, we extend to classifying multiple objects of interest. The CANUPO and RandLA-Net approaches were trained to identify submerged tyres and chairs. The results obtained from these approaches are comparable in terms of their overall accuracy levels and their F1 scores. The results of both models show good classification accuracy. This demonstrates good potential for applying these techniques in underwater cultural heritage preservation and awareness issues, landslide hazard mapping, disaster recovery expeditions,

underwater infrastructure inspections and the exploration of mineral resources on the seabed. Future research in this field should explore point cloud classifications in natural sub-aquatic environments and the building up of an extensive database of our underwater natural marine heritage.

6. Acknowledgements

Acknowledgements are all due to the South African National Research Foundation (NRF) for providing the financial support for this work. The National Lotteries Commission is acknowledged for funding the UKZN's Underwater Heritage Mapping and Documentation project. We acknowledge the reviewers for their contributions in improving upon this paper.

7. References

Alexander, C, Tansey, K, Kaduk, J, Holland, D & Tate, NJ 2011, 'An approach to the classification of airborne laser scanning point cloud data in an urban environment', *International Journal of Remote Sensing*, vol. 32 no. 24, pp. 9151-9169.

AVEVA, 2020, AVEVA Point Cloud Manager, Cambridge, United Kingdom, viewed 03 April 2024 <https://www.aveva.com/en/products/point-cloud-manager/>.

Boser, BE, Guyon, IM & Vapnik, VN 1992, 'A training algorithm for optimal margin classifiers', 1992 *Proceedings of the Fifth Annual Workshop on Computational Learning Theory, July 1992*, pp. 144-152.

Bowens, A (ed.), 2011, *Underwater archaeology: the NAS guide to principles and practice*. John Wiley & Sons.

Brodu, N & Lague, D 2012, '3D terrestrial LIDAR data classification of complex natural scenes using a multi-scale dimensionality criterion: Applications in Geomorphology', *ISPRS Journal of Photogrammetry and Remote Sensing*, vol. 68, pp. 121-134.

Coetzee, S & Singh, M 2021, 'Assessing the Relative Accuracy of the BlueView BV5000 Mechanical Scanning Sonar against Physical Measurements and Underwater Photogrammetry Models' [Unpublished Honours Thesis]. University of KwaZulu-Natal.

Cortes, C & Vapnik, V 1995, 'Support-vector networks', *Machine Learning*, vol. 20, pp. 273-297.

DSAC - Department of Sports, Arts and Culture, 2009, Launch of underwater cultural heritage projects, South Africa, viewed 23 June 2021, available from <http://www.dac.gov.za/content/launch-underwater-cultural-heritage-project>.

ESRI 2023, *Introduction to Deep Learning and Point Clouds*, ESRI, Midrand, South Africa, viewed 15 September 2023, available from <https://pro.arcgis.com/en/proapp/latest/help/data/las-dataset/introduction-to-deep-learning-and-point-clouds.htm>.

Foody, GM & Mathur, A 2004, 'Toward intelligent training of supervised image classifications: directing training data acquisition for SVM classification', *Remote Sensing of the Environment*, vol. 93 no. 1-2, pp. 107-117.

Forrest, TG 1994, 'From sender to receiver: propagation and environmental effects on acoustic signals', *American Zoologist*, vol. 34 no. 6, pp. 644-654.

Girardeau-Montaut, D 2023, CloudCompare (version 2.13beta) [GPL software], viewed 28 January 2024, available from <https://www.danielgm.net/cc/>.

Grilli, E, Özdemir, E & Remondino, F 2019, 'Application of machine and deep learning strategies for the classification of heritage point clouds', *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, vol. 42, pp. 447-454.

Hake, F, Lippmann, P. Alkhatib, H. Oettel, V & Neumann, I 2023, 'Automated damage detection for port structures using machine learning algorithms in heightfields', *Applied Geomatics*, pp. 1-9.

Himri, K, Ridao, P & Gracias, N 2019, '3D object recognition based on point clouds in an underwater environment with global descriptors: a survey', *Sensors*, vol. 19, no. 20, p.4451.

Hu, Q, Yang, B, Xie, L, Rosa, S, Guo, Y, Wang, Z, Trigoni, N & Markham, A 2020, 'Randla-net: Efficient semantic segmentation of large-scale point clouds', *2020 Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2020. IEEE pp. 11108-11117.

Hu, X, Zhang, Z, Duan, Y, Zhang, Y, Zhu & Long, H 2012, 'LIDAR photogrammetry and its data organization', *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, vol. 38, pp. 181-184.

Kogut, T & Slowik, A 2021, 'Classification of airborne laser bathymetry data using artificial neural networks', *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 14, pp. 1959-1966.

Langley, P & Simon, HA 1995, 'Applications of machine learning and rule induction', *Communications of the ACM*, vol. 38 no. 11, pp. 54-64.

Leica Geosystems 2017, Leica Cyclone REGISTER 360 PLUS - 3D Laser Scanning Point Cloud Registration Software, Switzerland, viewed 03 April 2024, available from <https://leica-geosystems.com/products/laser-scanners/software/leica-cyclone/leica-cyclone-register-360>.

Lesnikowski, N & Rush, B 2012, *Spool Piece Metrology Applications utilizing BV5000 3D Scanning Sonar*, Technical Report, BlueView Technologies: Seattle, WA, USA.

Li, Y, Bu, R, Sun, M, Wu, W, Di, X & Chen, B 2018, 'Pointcnn: Convolution on x-transformed points', in *Advances in Neural Information Processing Systems*, Montreal, January 2018, Canada, 2018.

Liarokapis, F, Kouřil, P, Agrafiotis, P, Demesticha, S, Chmelik, J & Skarlatos, D 2017, '3D modelling and mapping for virtual exploration of underwater archaeology assets', *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, vol. 42, pp. 425-431.

Lidar 360 2013, Lidar360 Point Cloud Post-processing Software, California, United States, viewed 05 April 2024, available from <https://www.greenvalleyintl.com//LiDAR360/>.

Long, J, Zhang, H & Zhao, J 2023, 'A Comprehensive Deep Learning-based Outlier Removal Method for Multibeam Bathymetric Point Clouds', *IEEE Transactions on Geoscience and Remote Sensing*, vol. 61, pp. 1-22.

Menna, F, Agrafiotis, P & Georgopoulos, A 2018, 'State of the art and applications in archaeological underwater 3D recording and mapping', *Journal of Cultural Heritage*, vol. 33, pp. 231-248.

Moisan, E, Charbonnier, P, Foucher, P, Grussenmeyer, P & Guillemin, S 2018, 'Evaluating a static multi-beam sonar scanner for 3D surveys in confined underwater environments', *Remote Sensing*, vol. 10, no. 9, p. 1395.

Ntuli, S & Forbes, A 2023, 'Classification of 3D UAS-SfM Point Clouds in the Urban Environment', *South African Journal of Geomatics*, vol. 12, no. 2, pp. 190-205.

Palomer, A, Ridao, P, Ribas, D & Vallicrosa, G 2015, 'Multi-beam terrain/object classification for underwater navigation correction', In: *OCEANS*, Genova, May 2015, pp. 1-5, IEEE, Genova, 2015.

Peeroo, U, Idrees, MO & Saeidi, V 2017, 'Building extraction for 3D city modelling using airborne laser scanning data and high-resolution aerial photo', *South African Journal of Geomatics*, vol. 6, no. 3, pp. 363-376.

Pepe, M, Alfio, VS, Costantino, D & Scaringi, D 2022, 'Data for 3D reconstruction and point cloud classification using machine learning in a cultural heritage environment', *Data in Brief*, vol. 42, p. 108250.

Pix4D SA, 2024, Pix4DMapper, Switzerland, viewed 05 April 2024, available from <https://www.pix4d.com/product/pix4dmapper-photogrammetry-software/>.

Pointly, 2024, Pointly, viewed 05 April 2024, available from <https://pointly.ai/>.

Qi, CR, Su, H, Mo, K & Guibas, LJ 2017, 'Pointnet: Deep learning on point sets for 3D classification and segmentation', *2017 Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2017, IEEE, pp. 652-660.

Qi, CR, Yi, L, Su, H & Guibas, LJ 2017, 'Pointnet++: Deep hierarchical feature learning on point sets in a metric space', in *Advances in Neural Information Processing Systems*, Long Beach, June 2017, USA,2017.

Richter, R, Behrens, M & Döllner, J 2013, 'Object class segmentation of massive 3D point clouds of urban areas using point cloud topology', *International Journal of Remote Sensing*, vol. 34 no. 23, pp. 8408-8424.

Rusu, RB & Cousins, S 2011, '3D is here: Point cloud library (pcl)', *2011 IEEE International Conference on Robotics and Automation, 2011, IEEE,* pp. 1-4.

Sung, M, Kim, J, Cho, H, Lee, M & Yu, SC 2020, 'Underwater-sonar-image-based 3D point cloud reconstruction for high data utilization and object classification using a neural network', *Electronics*, vol. 9 no. 11, p. 1763.

Teledyne Marine, 2018, Teledyne Marine, Slangerup, Denmark, viewed 1 April 2023, available from <http://www.teledynemarine.com/Blueview/>.

Teledyne Marine, 2020, Teledyne Marine, Slangerup, Denmark, viewed 5 April 2023, available from <http://www.teledynemarine.com/Lists/Downloads/Blueview/Blueview BV5000 Product leaflet.pdf>.

The Cross Product, 2020, TheCrossProduct - LiDAR Point Cloud Processing Software, viewed 05 April 2024, available from <https://www.thecrossproduct.com>.

The HDF Group, 2023, HDF5 - Hierarchical Data Format, viewed 29 March 2024: available from https://www.hdfgroup.org/solutions/hdf5/.

Tsai, CM, Lai, YH, Sun, YD, Chung, YJ & Shortis, M 2021, 'Multi-dimensional underwater point cloud detection based on deep learning', *Sensors*, vol. 21 no. 3, p. 884.

UNESCO – United Nations Educational, Scientific and Cultural Organisation, 2015, *Operational Guidelines for the Convention on the Protection of the Underwater Cultural Heritage*, UNESCO, United Nations, viewed 4 April 2023, available from <unesdoc.unesco.org/images/0023/002341/234177E.pdf>.

Van Genechten, B 2008, *Theory and practice in Terrestrial Laser Scanning: Training material based on practical applications*, Universidad Politecnica de Valencia Editorial, Valencia, Spain.

Wang, H & Raj, B 2017, 'On the origin of deep learning', *arXiv preprint arXiv:1702.07800.*

Weinmann, M, Jutzi, B, Hinz, S & Mallet, C 2015, 'Semantic point cloud interpretation based on optimal neighborhoods, relevant features and efficient classifiers', *ISPRS Journal of Photogrammetry and Remote Sensing*, vol. 105, pp. 286-304.

Westoby, MJ, Brasington, J, Glasser, NF, Hambrey, MJ & Reynolds, JM 2012, 'Structure-from-Motion' photogrammetry: A low-cost, effective tool for Geoscience applications', *Geomorphology*, vol. 179, pp. 300-314.

Willmott, CJ 1982, 'Some comments on the evaluation of model performance', *Bulletin of the American Meteorological Society*, vol. 63 no. 11, pp. 1309-1313.

Wu, W, Qi, Z & Fuxin, L 2019. 'Pointconv: Deep convolutional networks on 3D point clouds'. *2019 Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, IEEE, 2019, pp. 9621-9630.

Zhang, W, Qi, J, Wan, P, Wang, H, Xie, D, Wang, X. & Yan, G 2016, 'An easy-to-use airborne LiDAR data filtering method based on cloth simulation', *Remote Sensing*, vol. no. 86, p. 501.