

UTILIZING DECISION TREE-BASED THRESHOLDS IN RECONSTRUCTING TRAINING DATASETS FOR SINGLE-CLASSIFICATION BINARY SUPPORT VECTOR MACHINE PROBLEM FROM POINT CLOUD DATA

Shydel M. Sarte¹, Alejandro H. Ballado, Jr.^{1,2}, Felicito S. Caluyo¹, Ramon G. Garcia^{1,2}, Glenn V. Magwili^{1,2}, Jennifer C. Dela Cruz^{1,2}, Sarah Alma P. Bentir¹

¹School of Graduate Studies, Mapúa University

²School of Electrical, Electronics and Computer Engineering, Mapúa University
Muralla St., Intramuros 1002 Manila, Philippines

shydelsarte@gmail.com

ahballado@mapua.edu.ph

fscaluyo@mapua.edu.ph

rggarcia@mapua.edu.ph

gvmagwili@mapua.edu.ph

jcdelacruz@mapua.edu.ph

sapbentir@gmail.com

KEY WORDS: Cocos nucifera, LiDAR, Canopy Height Model, Support Vector Machine, Decision Tree

ABSTRACT: In some cases of agricultural features extraction, the binary Support Vector Machine (SVM) classification problem only requires the objective of mapping a single class across a relatively large geographic area without the need to identify other land use and land cover classes. While the conduct of ground trothing to collect field data samples is a requirement to achieve high accuracy levels, sometimes it is impractical considering the labor and cost of conducting field surveys that necessitates single species only. As such, the Cocos Nucifera or coconut, which is an agricultural class that is visually discernible through LiDAR-derived Canopy Height Model (CHM), may be classified without the need for in situ data collection. This is done by exploiting the threshold values calculated via Decision Tree (DT) algorithm in reconstructing datasets. In comparison, classification outputs from in situ training samples and from the DT-derived samples achieve similar accuracy levels, hence this study introduces a classification methodology that eliminates the need for field data gathering and manual ‘training data’ selection in mapping coconut species. A single-classification binary SVM has been implemented using LiDAR-derived CHM, utilizing only the elevation information contained in the point cloud data, with a grid resolution of one meter. The results suggest that automatic selection of samples is tolerable given that a representative calibration site is identified.

1. INTRODUCTION

The Phil-LiDAR 2 Program, also known as the Nationwide Detailed Resources Assessment using LiDAR, is a three-year program funded by the Department of Science and Technology (Blanco, Tamondong, Perez, Ang, & Paringit, 2015). The program, which started July 2014, aims to use the LiDAR datasets in order to extract various natural resources including agricultural, coastal, hydrological, forest and renewable energy. The acquired data achieves an equivalent 1 meter grid resolution and have been used to map natural resources in the Philippines.

LiDAR have so much applications in agriculture, one of which is this research endeavor which focuses on the objective of developing an alternative methodology for mapping coconut resources via linear Support Vector Machine classification. In particular, this study aims at generating training datasets without the need to conduct intensive field data gathering. This is implemented through the decision tree algorithm.

1.1. Resources Mapping

Geographic information systems (GIS) technology was originally conceptualized to advance digital processing of spatial data. As modern natural resource management deals with temporal and spatial information, GIS has been implemented in the analysis, modelling and visualization datasets (Fuller, Wang, Gross, & Berry, 2007). In particular, GIS applications, such as resource mapping, land use and land cover mapping, and forest parameter estimations, have introduced more advanced means in the management of natural resources.

In their introspection on the variability of the natural resources, Blanco, et al. raised the need to generate a “detailed and up-to-date inventory of the natural resources” in the Philippines (2015) in the form of maps and in the aid Geographic Information Systems (GIS) applications. Therefore, the nationwide detailed resources assessment program known as Phil-LiDAR 2 has been implemented to complement current government programs by producing high-resolution natural resources maps in the Philippines.

1.2. LiDAR technology

RS has been supplementing field measurements in the recent years (Wu, et al., 2016). In addition, with the availability of high resolution image data and through the incorporation of more advanced methodologies, accurate segmentation and classification of features can be achieved. The transition from the conventional pixel-based approach to the object-based image analysis (OBIA) in conducting image processing has also added more capabilities for various approaches and applications.

There are a number of remote sensing images in use today that enhance GIS products. Satellite imageries, for instance, are now accessible in the public. The use of multispectral and hyperspectral images paved for more capabilities particularly in land use and land cover mapping. With the advent of LiDAR technology, more accurate maps have been produced due to the elevation information contained in the data.

LiDAR stands for Light Detection and Ranging. As simply defined by the National Oceanic and Atmospheric Administration (NOAA), LiDAR is a remote sensing method that is “used to examine the surface of the Earth” using pulsed laser. LiDAR technology is capable of collecting data of high resolution and high accuracy on a broad yet scalable coverage in a cost-effective means. According to the National Research Council, it is “the most cost-effective technology to acquire elevation information over large regions to support floodplain mapping to FEMA accuracy standards (2007).”

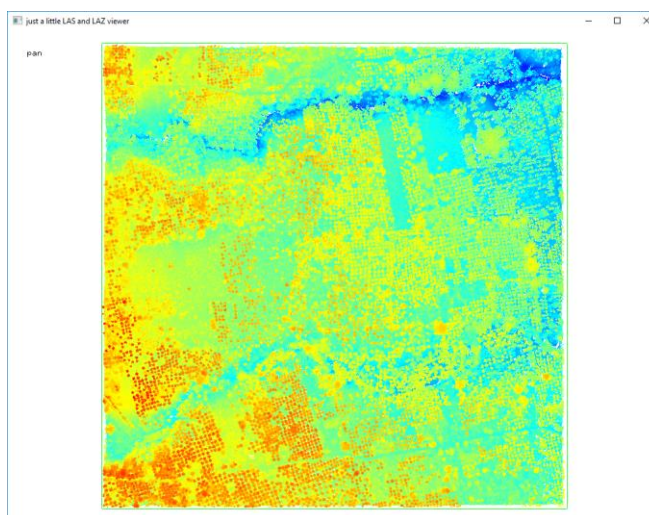


Figure 1. Depiction of LiDAR data using lastools software

LiDAR data (see depiction via Figure 1) contain regularly-spaced points forming a “cloud” with each point containing horizontal and vertical position values (x, y and z) with respect to some horizontal and vertical datum. LiDAR data also include information about return intensity and return number. LiDAR, with its high resolution, can be used in accurately deducing geometric properties from delineated tree crown structures (Wu, et al., 2016).

1.3 Object-Based Image Analysis

Object-based image analysis (OBIA) has been introduced to improve image processing and eliminate the speckles that are inherent in the pixel-based process. Instead of involving pixel classification, OBIA works on objects, which are outlined regions with homogeneous properties (Yadav, Rizvi, & Kadam, 2015), with segmentation being the heart of the process. Segmentation refers to the method of generating object outlines. It refers to the creation of objects that are an aggregation of pixels with similar features. Figure 2 shows delineation of objects applied in an elevation layer. Pixels of similar elevation values aggregate together, forming the objects.

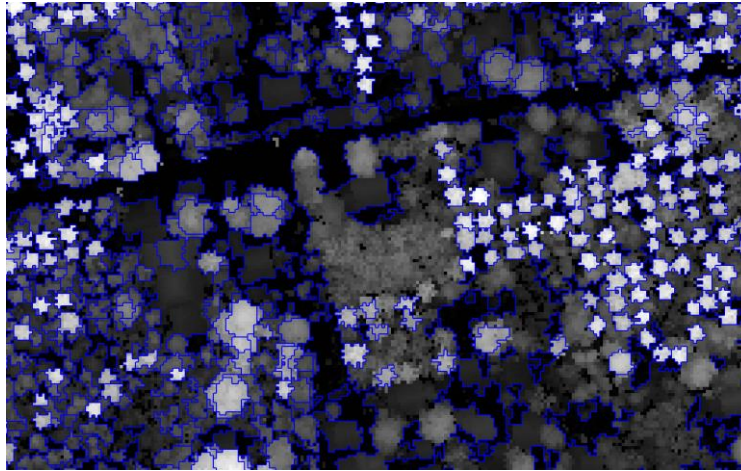


Figure 2. Image scene shows delineated objects, viewed via eCognition

Kumar, et al. discussed that image segmentation can be categorized depending on the approach, such as thresholding, region-based, edge-based, ANN-based, PDE-based and Fuzzy-based (2016).

Landcover mapping is fundamental in utilizing RS data for natural resources management. While segmentation is the core of the OBIA process, derivation of information from RS images relies on “robust classification methods” (Maulik & Chakraborty, 2017). Image classification involves “image pre-processing, the detection and extraction of an object, feature extraction, selection of training samples, post-classification processing and accuracy assessment.”

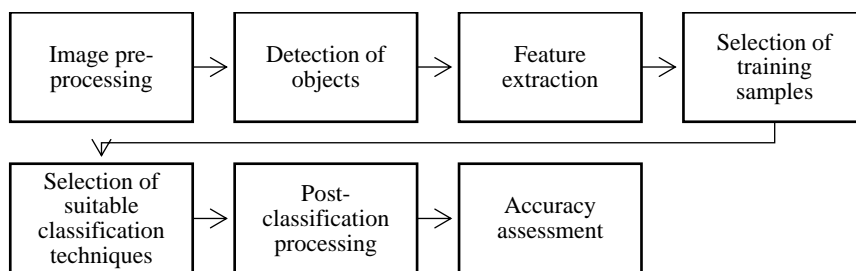


Figure 3. Major steps in image classification (Maulik & Chakraborty, 2017)

Machine learning plays an important role in the context of image classification. In RS application, machine learning algorithms eliminate the task of manual selection of land cover classes. Identification of suitable machine learning algorithms considers effectiveness and efficiency. Machine learning can be unsupervised and supervised. Advanced classification procedures use supervised methods in which the training sets of examples are used in identifying class boundaries. Common supervised learning methods are nearest neighbor, decision tree and Support Vector Machine (SVM) algorithms.

The Support Vector Machine (SVM) classifier is one of the most commonly preferred algorithms in RS applications due to its “good classification performance with high-dimensional data” (Baldeck & Asner, 2015) and it is even labeled as the top-performing classifier algorithm when dealing with hyperspectral data in recent studies. While SVM methods were originally tested in the classification of as much classes as possible in an image scene, research objectives that deal with one or a few classes are not uncommon.

2. METHODOLOGY

2.1. Study Area and Materials

Flight mission “Laguna_Blk18VWs_20140425” crosses the municipality of San Antonio in the province of Quezon. 16 LiDAR tiles are selected for processing, with each tile covering a 1-by-1-km grid. These 16 tiles completely cover barangays Niing, Bagong Niing and Poblacion. The province of Quezon has been selected as it is one of the top producers of coconuts in the Philippines. The municipality of San Antonio is picked due to considerations on the variability of features, safety and accessibility. 16 tiles have been used to fit in the study time frame.

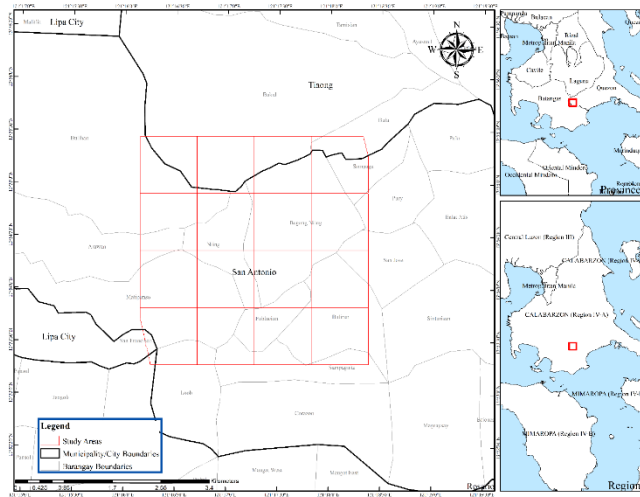


Figure 4. Selection of LiDAR data tiles in the study area

The raster elevation layers are shown in Figure 5 for all the study sites selected inside the municipality of San Antonio in the province of Quezon. The general workflow of this study is applied separately to all the 16 study areas, with each site having its own training and validation samples, and separate accuracy assessment matrices.

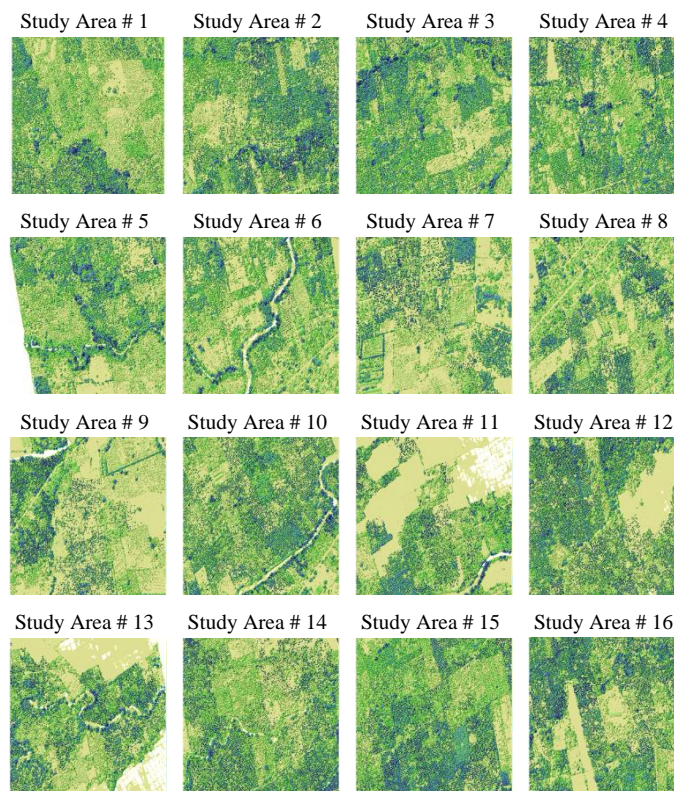


Figure 5. Depiction of 16 study sites in San Antonio Quezon

2.2. Methodological Workflow

The overall workflow starts with the selection of the LiDAR datasets. Then, the elevation values are rasterized as a Canopy Height Model (CHM) that provides height values of features above the ground. The raster datasets are loaded in the eCognition software in the implementation of Object-Based Image Analysis. First, a series of segmentation procedures are executed until the individual trees are delineated. Then, the ideal classification layers are generated using manual classification. The other feature layers are also generated prior to the training of the decision tree algorithm. When the feature layers and the classified images are created, the decision tree algorithm is applied to extract the separability threshold for the coconuts and other classes. These threshold values are used to generate training data sets for the implementation of the Support Vector Machine classifier algorithm. After executing the

SVM classifier, accuracy assessment is performed on all the outputs of the 16 eCognition projects. Figure 6 summarizes the overall workflow.

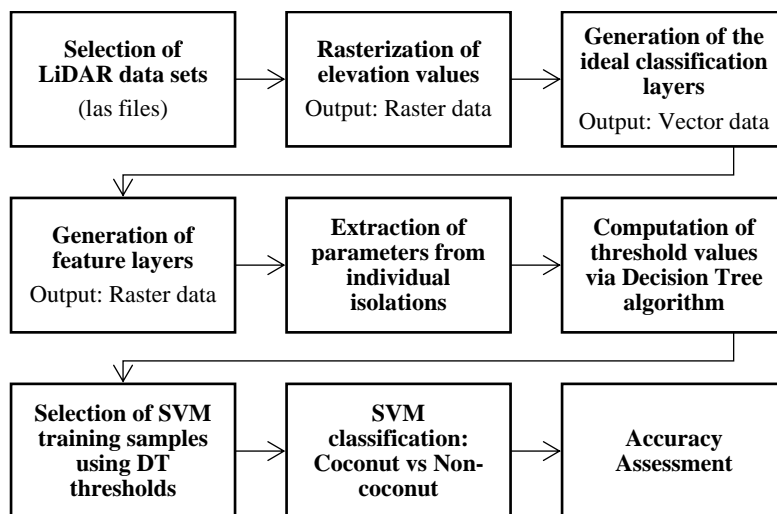


Figure 6. Implemented methodological workflow of the study

The classes identified through the OBIA process are summarized in Table 1 with their corresponding descriptions.

Table 1. Classes identified in the OBIA process

| Parent Class | Class | Sub-Class | Description |
|-----------------|--------------|---------------|--|
| Class_Ground | Class_Ground | Class_Ground | Objects on the ground |
| Class_NonGround | Lower | Lower | Objects above the ground but below coconut height threshold values. This class should not include coconut trees. |
| Class_NonGround | Upper | Upper | Objects above the ground that may include coconut trees, tall structures and other vegetation classes |
| Class_NonGround | Upper | Class_Coconut | Objects that are identified as coconut trees |
| Class_NonGround | Upper | Diff_Coconut | Non-Coconut objects classified as class_Coconut |

Essentially, two sets of classification images, each consists of 16 study sites, are produced in this study: (1) SVM classified images using field data, (2) SVM classified image using training samples generated by applying the DT thresholds. For every study area, samples are gathered in two 100m-by-100m plots (see Figure 7). DT-based generation of samples is also implemented in the same 100m-by-100m grids where the field training samples are located. The DT-based samples are classified according to the thresholds identified by the decision trees. Generally, the mean values are applied as thresholds as long as they represent at least half of the 16 study sites.

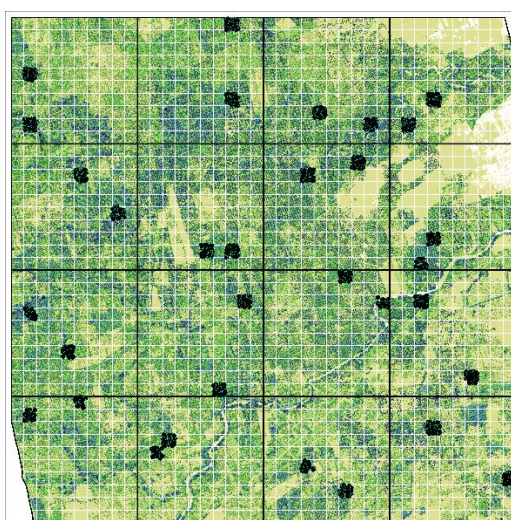


Figure 7. Training samples in 100m-by-100m grids

3. RESULTS AND DISCUSSION

3.1. Manually classified layers

All of the manually generated coconut tree objects are exported as a shapefile for all the 16 study sites. Descriptive statistics are calculated for the feature layers elevation, intensity, area and elliptic fit. The exported shapefile of manually classified coconut objects for study area no. 12 is shown in Figure 8.

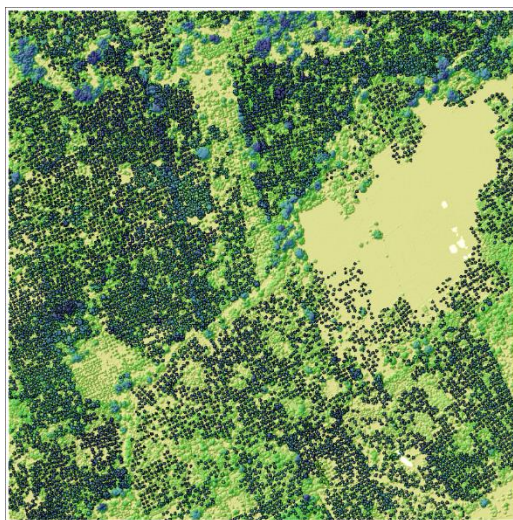


Figure 8. Delineated coconut classes in Study Area 12

Table 2 summarizes some descriptive statistics of the feature layers in terms of range, mean and standard deviation.

Table 2. Summary of feature statistics for ideal classes

| Feature | Range | Mean | Std Dev |
|--------------|-----------|-------------|-------------|
| Area | 6.250734 | 24.44261694 | 2.258818613 |
| Elevation | 5.734089 | 15.8615875 | 1.888205247 |
| Elliptic Fit | 0.04268 | 0.744761688 | 0.012538624 |
| Intensity | 31.304401 | 7.105228375 | 10.18200848 |

3.2. Decision Trees

Decision trees are generated for all the 16 study areas using the feature values of the manually classified layers. One of the decision trees is shown in Figure 9, which represents study area no. 7. The tabulation of threshold values are based on the immediate nodes from the class_Coconut, which are 1.5 and 9.976 here for area and elevation respectively.

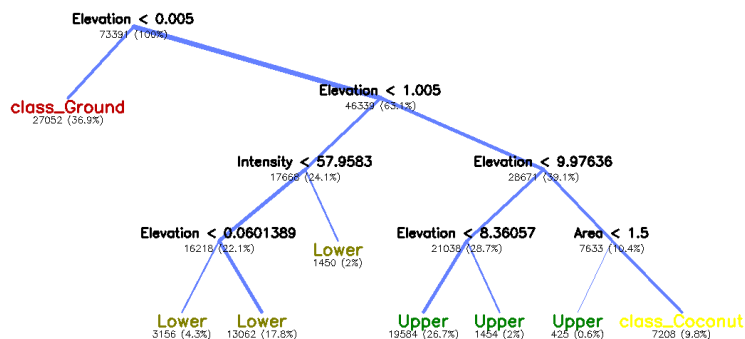


Figure 9. Decision tree and thresholds generated in Study Area No. 7

The tabulation of decision tree thresholds of 16 study sites is provided in Table 3. We can immediately notice the convergence of the values. The mean of the elevation thresholds can then be used as the overall threshold value of all study areas. Surprisingly, the area threshold settles at an exact value.

Table 3 Decision tree thresholds of feature layers for classCoconut

| Study Area | Source LAS number | Elevation (m) | Intensity | Area (Pxl) | Elliptic Fit |
|------------|-------------------|---------------|-----------|------------|--------------|
| 1 | Pt000222 | 9.99857 | N/A | 1.5 | N/A |
| 2 | Pt000224 | 9.99588 | N/A | 1.5 | N/A |
| 3 | Pt000226 | 9.99014 | N/A | N/A | N/A |
| 4 | Pt000227 | 10.0005 | N/A | 1.5 | N/A |
| 5 | Pt000228 | 8.84086 | N/A | N/A | N/A |
| 6 | Pt000229 | 9.89558 | N/A | N/A | N/A |
| 7 | Pt000264 | 9.97636 | N/A | 1.5 | N/A |
| 8 | Pt000265 | 10.0022 | N/A | 1.5 | N/A |
| 9 | Pt000266 | 10.0013 | N/A | 1.5 | N/A |
| 10 | Pt000267 | 10.0023 | N/A | 1.5 | N/A |
| 11 | Pt000268 | 10.0005 | N/A | 1.5 | N/A |
| 12 | Pt000269 | 10.0031 | N/A | 1.5 | N/A |
| 13 | Pt000270 | 10.0191 | N/A | 1.5 | N/A |
| 14 | Pt000271 | 9.99754 | N/A | 1.5 | N/A |
| 15 | Pt000272 | 9.99789 | N/A | 1.5 | N/A |
| 16 | Pt000286 | 10.0004 | N/A | 1.5 | N/A |

Immediately, we can apply the thresholds to classify the coconut objects. The result of this operation, known as classification using threshold values, can be compared to the ideal classification layers from which we can identify which non-coconut classes are misclassified by the threshold-approach classification. The difference in coconut classes between the threshold-based classification and the ideal classification layers reveals non-coconut objects that are prone to be classified as coconuts, and thus labeled as diff_Coconut. Then, we are applying the decision tree algorithm among class_Coconut and diff_Coconut to determine their separability thresholds. The generated decision tree for study area no. 7 is provided in Figure 10.

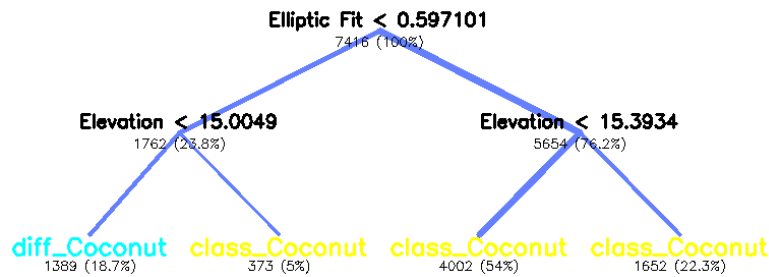


Figure 10. Decision tree and thresholds generated in Study Area No. 7

Finally, we can see the decision tree thresholds that separated the class Coconut and the diff_Coconut classes and are tabulated in Table 4.

Table 4 Decision tree thresholds of feature layers for diff_Coconut classes

| Study Area | Source LAS number | Elevation | Intensity | Area | Elliptic Fit |
|------------|-------------------|--------------------|-----------|------|--------------|
| 1 | Pt000222 | 9.92339 20.795 | | | 0.593712 |
| 2 | Pt000224 | 10.1755 14.9946 | | | 0.593976 |
| 3 | Pt000226 | 15.0045 | 28.467 | | 0.597297 |
| 4 | Pt000227 | 15.0544 | | 52.5 | 0.593907 |
| 5 | Pt000228 | 9.92494 15.9471 | | | 0.594079 |
| 6 | Pt000229 | 15.0173 15.8153 | | | 0.596296 |
| 7 | Pt000264 | 15.0049 15.3934 | | | 0.597101 |

| | | | | | |
|----|----------|---------|---------|------|----------|
| 8 | Pt000265 | 15.0006 | 4.53814 | | 0.596875 |
| 9 | Pt000266 | 15.0385 | | 37.5 | 0.595918 |
| 10 | Pt000267 | 14.9952 | 3.03391 | | 0.596875 |
| 11 | Pt000268 | 15.0423 | | | 0.597101 |
| 12 | Pt000269 | 16.3846 | 6.29865 | | 0.557439 |
| 13 | Pt000270 | 14.9966 | | 37.5 | 0.59661 |
| 14 | Pt000271 | 15.6569 | 1.30268 | | 0.927249 |
| 15 | Pt000272 | 15.0054 | | 56.5 | 0.597619 |
| 16 | Pt000286 | 15.0182 | | 41.5 | 0.59798 |

3.3. Accuracy Assessment of SVM Classified Image Using Field Data Samples

Table 5. Overall accuracies of the classified layers using field-acquired training datasets

| Study Area | LAS No. | Class | Producer | User | Hellden | Short | KIA Per Class | Overall Accuracy | KIA |
|------------|----------|-------------|----------|--------|---------|--------|---------------|------------------|--------|
| 1 | Pt000222 | Coconut | 1.0000 | 0.9427 | 0.9705 | 0.9427 | 1.0000 | 0.9623 | 0.9184 |
| | | Not Coconut | 0.9008 | 1.0000 | 0.9478 | 0.9008 | 0.8492 | | |
| 2 | Pt000224 | Coconut | 0.8790 | 0.9645 | 0.9198 | 0.8515 | 0.7267 | 0.9062 | 0.8075 |
| | | Not Coconut | 0.9490 | 0.8328 | 0.8871 | 0.7971 | 0.9085 | | |
| 3 | Pt000226 | Coconut | 0.7896 | 0.8788 | 0.8318 | 0.7121 | 0.6396 | 0.8521 | 0.7004 |
| | | Not Coconut | 0.9060 | 0.8330 | 0.8680 | 0.7667 | 0.7742 | | |
| 4 | Pt000227 | Coconut | 0.8894 | 0.9874 | 0.9358 | 0.8794 | 0.7647 | 0.9282 | 0.8549 |
| | | Not Coconut | 0.9837 | 0.8615 | 0.9186 | 0.8494 | 0.9693 | | |
| 5 | Pt000228 | Coconut | 0.8739 | 0.9122 | 0.8927 | 0.8061 | 0.7666 | 0.8992 | 0.7976 |
| | | Not Coconut | 0.9224 | 0.8880 | 0.9049 | 0.8263 | 0.8313 | | |
| 6 | Pt000229 | Coconut | 0.9195 | 0.8946 | 0.9069 | 0.8296 | 0.8613 | 0.9229 | 0.8411 |
| | | Not Coconut | 0.9252 | 0.9434 | 0.9342 | 0.8765 | 0.8218 | | |
| 7 | Pt000264 | Coconut | 0.8797 | 1.0000 | 0.9360 | 0.8797 | 0.7324 | 0.9247 | 0.8455 |
| | | Not Coconut | 1.0000 | 0.8325 | 0.9086 | 0.8325 | 1.0000 | | |
| 8 | Pt000265 | Coconut | 0.8828 | 0.9685 | 0.9237 | 0.8581 | 0.8037 | 0.9355 | 0.8681 |
| | | Not Coconut | 0.9773 | 0.9133 | 0.9442 | 0.8943 | 0.9436 | | |
| 9 | Pt000266 | Coconut | 0.9504 | 0.8687 | 0.9077 | 0.8311 | 0.8833 | 0.8985 | 0.7954 |
| | | Not Coconut | 0.8410 | 0.9387 | 0.8872 | 0.7973 | 0.7234 | | |
| 10 | Pt000267 | Coconut | 0.8974 | 0.9571 | 0.9263 | 0.8627 | 0.7867 | 0.9209 | 0.8412 |
| | | Not Coconut | 0.9501 | 0.8819 | 0.9148 | 0.8429 | 0.9039 | | |
| 11 | Pt000268 | Coconut | 0.9202 | 0.8696 | 0.8942 | 0.8086 | 0.8539 | 0.9066 | 0.8107 |
| | | Not Coconut | 0.8963 | 0.9373 | 0.9164 | 0.8457 | 0.7716 | | |
| 12 | Pt000269 | Coconut | 0.9108 | 0.9666 | 0.9378 | 0.8830 | 0.7658 | 0.9207 | 0.8286 |
| | | Not Coconut | 0.9397 | 0.8462 | 0.8905 | 0.8026 | 0.9026 | | |
| 13 | Pt000270 | Coconut | 0.9517 | 0.9694 | 0.9605 | 0.9239 | 0.8877 | 0.9545 | 0.9070 |
| | | Not Coconut | 0.9585 | 0.9348 | 0.9465 | 0.8985 | 0.9272 | | |
| 14 | Pt000271 | Coconut | 0.9078 | 0.9562 | 0.9314 | 0.8716 | 0.8025 | 0.9249 | 0.8485 |
| | | Not Coconut | 0.9468 | 0.8891 | 0.9170 | 0.8467 | 0.9001 | | |
| 15 | Pt000272 | Coconut | 0.9490 | 0.9739 | 0.9613 | 0.9255 | 0.8985 | 0.9609 | 0.9219 |
| | | Not Coconut | 0.9734 | 0.9481 | 0.9606 | 0.9242 | 0.9465 | | |
| 16 | Pt000286 | Coconut | 0.9432 | 0.9836 | 0.9630 | 0.9286 | 0.8857 | 0.9619 | 0.9238 |
| | | Not Coconut | 0.9826 | 0.9400 | 0.9608 | 0.9246 | 0.9654 | | |

Execution of SVM classification using the field coconut samples achieves accuracy levels of around 92.4% with a minimum accuracy level of 85%. This is already sufficient given that we are not implementing any refinement methods that are outside the SVM classifier. The statistics of the overall accuracies are provided in Table 6.

Table 6. Descriptive statistics of the overall accuracies of 16 sites

| Statistics | Value |
|----------------|-------------|
| Mean | 0.923752838 |
| Standard Error | 0.007199131 |
| Median | 0.923795402 |
| Mode | N/A |

| | |
|--------------------|-------------|
| Standard Deviation | 0.028796523 |
| Sample Variance | 0.00082924 |
| Kurtosis | 1.294439862 |
| Skewness | -0.68500576 |
| Range | 0.110241767 |
| Minimum | 0.85206422 |
| Maximum | 0.962305987 |

3.4. Accuracy Assessment of SVM Classified Images Through DT-Generated Samples

Table 7. Overall accuracies of the classified layers using threshold-based training datasets

| Study Area | LAS No. | Class | Producer | User | Hellden | Short | KIA Per Class | Overall Accuracy | KIA |
|------------|----------|-------------|----------|--------|---------|--------|---------------|------------------|--------|
| 1 | Pt000222 | Coconut | 1.0000 | 0.9012 | 0.9480 | 0.9012 | 1.0000 | 0.9320 | 0.8505 |
| | | Not Coconut | 0.8210 | 1.000 | 0.9017 | 0.8210 | 0.7399 | | |
| 2 | Pt000224 | Coconut | 0.9633 | 0.9781 | 0.9706 | 0.9429 | 0.9077 | 0.9643 | 0.9252 |
| | | Not Coconut | 0.9660 | 0.9435 | 0.9546 | 0.9132 | 0.9435 | | |
| 3 | Pt000226 | Coconut | 0.9010 | 0.9430 | 0.9215 | 0.8545 | 0.8224 | 0.9289 | 0.8566 |
| | | Not Coconut | 0.9530 | 0.9177 | 0.9350 | 0.8780 | 0.8938 | | |
| 4 | Pt000227 | Coconut | 0.9350 | 0.9880 | 0.9607 | 0.9244 | 0.8532 | 0.9550 | 0.9082 |
| | | Not Coconut | 0.9837 | 0.9136 | 0.9474 | 0.9000 | 0.9708 | | |
| 5 | Pt000228 | Coconut | 0.9137 | 0.9157 | 0.9147 | 0.8429 | 0.8345 | 0.9183 | 0.8362 |
| | | Not Coconut | 0.9224 | 0.9206 | 0.9215 | 0.8544 | 0.8380 | | |
| 6 | Pt000229 | Coconut | 0.9195 | 0.8946 | 0.9069 | 0.8296 | 0.8613 | 0.9229 | 0.8411 |
| | | Not Coconut | 0.9252 | 0.9434 | 0.9342 | 0.8765 | 0.8218 | | |
| 7 | Pt000264 | Coconut | 0.9458 | 0.9522 | 0.9490 | 0.9029 | 0.8567 | 0.9364 | 0.8645 |
| | | Not Coconut | 0.9207 | 0.9104 | 0.9155 | 0.8442 | 0.8724 | | |
| 8 | Pt000265 | Coconut | 1.0000 | 0.8989 | 0.9468 | 0.8989 | 1.0000 | 0.9503 | 0.9004 |
| | | Not Coconut | 0.9110 | 1.0000 | 0.9534 | 0.9110 | 0.8189 | | |
| 9 | Pt000266 | Coconut | 0.9504 | 0.9455 | 0.9479 | 0.9010 | 0.8949 | 0.9451 | 0.8899 |
| | | Not Coconut | 0.9393 | 0.9448 | 0.9420 | 0.8904 | 0.8851 | | |
| 10 | Pt000267 | Coconut | 0.8974 | 1.0000 | 0.9459 | 0.8974 | 0.7961 | 0.9432 | 0.8865 |
| | | Not Coconut | 1.0000 | 0.8872 | 0.9402 | 0.8872 | 1.0000 | | |
| 11 | Pt000268 | Coconut | 0.9202 | 0.8696 | 0.8942 | 0.8086 | 0.8539 | 0.9066 | 0.8107 |
| | | Not Coconut | 0.8963 | 0.9373 | 0.9164 | 0.8457 | 0.7716 | | |
| 12 | Pt000269 | Coconut | 0.9685 | 0.9685 | 0.9685 | 0.9389 | 0.9082 | 0.9586 | 0.9082 |
| | | Not Coconut | 0.9397 | 0.9397 | 0.9397 | 0.8863 | 0.9082 | | |
| 13 | Pt000270 | Coconut | 0.9517 | 0.9581 | 0.9548 | 0.9136 | 0.8859 | 0.9478 | 0.8929 |
| | | Not Coconut | 0.9424 | 0.9338 | 0.9381 | 0.8834 | 0.9001 | | |
| 14 | Pt000271 | Coconut | 0.9584 | 0.9584 | 0.9584 | 0.9202 | 0.9052 | 0.9533 | 0.9052 |
| | | Not Coconut | 0.9468 | 0.9468 | 0.9468 | 0.8989 | 0.9052 | | |
| 15 | Pt000272 | Coconut | 0.9666 | 0.9743 | 0.9704 | 0.9425 | 0.9322 | 0.9699 | 0.9398 |
| | | Not Coconut | 0.9734 | 0.9653 | 0.9693 | 0.9405 | 0.9475 | | |
| 16 | Pt000286 | Coconut | 0.9432 | 0.9507 | 0.9470 | 0.8992 | 0.8815 | 0.9445 | 0.8888 |
| | | Not Coconut | 0.9460 | 0.9378 | 0.9419 | 0.8902 | 0.8963 | | |

Similarly, SVM classification is implemented but this time the training samples are generated using the DT thresholds calculated. Accuracy levels on average are around 94.2%, which is better compared to the results of the field data samples train. Moreover, the minimum accuracy level determined is 90.7%, a significant difference between the SVM results of the field data sample train.

Table 8. Descriptive statistics of the overall accuracies of 16 sites

| Statistics | Value |
|--------------------|--------------|
| Mean | 0.94232205 |
| Standard Error | 0.004309096 |
| Median | 0.94483337 |
| Mode | N/A |
| Standard Deviation | 0.017236383 |
| Sample Variance | 0.000297093 |
| Kurtosis | -0.227242611 |

| | |
|----------|--------------|
| Skewness | -0.455747085 |
| Range | 0.063315276 |
| Minimum | 0.906578947 |
| Maximum | 0.969894223 |

We can see that the overall accuracies achieved using DT-generated samples are less scattered with a standard deviation of 0.0172, or 1.72%. The relatively small levels of kurtosis and skewness indirectly supports agreement of the accuracies in terms of the distribution curve. Moreover, an accuracy of 96.989% is high enough to not consider refinement methods in improving the results of the classification. The KIA are also in agreement as shown in the table and we are observing sufficiently high levels on producer and user accuracies considering the 16 study sites.

4. CONCLUSION

In this research, DT-generated samples of coconut trees are used as SVM training data sets and are able to achieve similar accuracy levels without the need for actual field surveys. This is as long as the selected sample sites, where the DT-based samples are calculated, provide representative sample data sets.

While accuracies using field data samples are computed around 92.4%, the mean overall accuracy in the classification of 16 study sites via DT-generated SVM training samples is as much as 94.2%. In addition to the improvement seen in the accuracy levels, the cost of conducting field surveys is eliminated. This suggests that there are several thresholds that can be used to generate samples of coconut classes and the classification can still achieve high accuracy levels even without field data surveys conducted.

ACKNOWLEDGEMENT: The authors acknowledge the assistances of the Department of Science and Technology (DOST) Engineering Research and Development for Technology (ERDT), the Philippine Council for Industry, Energy and Emerging Technology Research and Development (PCIEERD), the Mapua Phil-LiDAR 2 and the Mapua University School of Graduate Studies (GS), in the fulfilment of this research endeavor. Data and software programs utilized for this paper are provided by the Phil-LiDAR 2 Program: Nationwide Detailed Resources Assessment Using LiDAR and other Remotely Sensed Data, headed by Dr. Ariel C. Blanco of University of the Philippines Diliman.

REFERENCES

- Baldeck, C. A., & Asner, G. P. (2015). Single-species detection with airborne imaging spectroscopy data : a comparison of support vector techniques. *IEEE Journal on Selected Topics in Applied Earth Observations and Remote Sensing*, 2501-2512.
- Blanco, A. C., Tamondong, A., Perez, A., Ang, M., & Paringit, E. (2015). The Phil-LiDAR 2 Program: National Resource Inventory of the Philippines Using LiDAR and Other Remotely Sensed Data. *36th International Symposium on Remote Sensing of Environment*. Berlin, Germany.
- FAO, F. a. (2017, May 20). *Climate Change*. Retrieved from Food and Agriculture Organization of the United Nations: <http://www.fao.org/climate-change/en/>
- Fuller, M. M., Wang, D., Gross, L. J., & Berry, M. W. (2007). Computational science for resource management. *Computing in Science and Engineering*, 40-48.
- Gulbe, L. (2015). Identification and delineation of individual tree crowns using lidar and multispectral data fusion. *2015 IEEE International Geoscience and Remote Sensing Symposium (IGARSS)* (pp. 3294-3297). Milan, Italy: IEEE.
- Maulik, U., & Chakraborty, D. (2017). Remote sensing image classification - a survey of support-vector-machine-based advanced techniques. *IEEE Geoscience and Remote Sensing Magazine*, 33-52.
- Wu, B., Yu, B., Wu, Q., Huang, Y., Chen, Z., & Wu, J. (2016). Individual tree crown delineation using localized contour tree method and airborne LiDAR data in coniferous forests. *International Journal of Applied Earth Observation and Geoinformation*, 82-94.
- Yadav, S., Rizvi, I., & Kadam, S. (2015). Urban tree canopy detection using Object-Based Image Analysis for Very High Resolution Satellite Images: a literature review. *2015 International Conference on Technologies for Sustainable Development* (pp. 1-6). Mumbai, India: IEEE.