One shot learning for dictionary classification of tropical rainforest species

Sng Li Wen, Aaron
Nanyang Technological University

Dr. Ji-jon Sit
Nanyang Technological University

Abstract - Tropical rainforests are dubbed the lungs of the world. With the advent of rapid urbanisation and climate change, there have been calls for conservatory efforts for our rainforests. Rainforest monitoring has advanced in recent years, but the task is arduous. The process of tracking of individual trees spans across vast expanses of land. These conservatory projects often require significant time and effort. In recent years, the technology for aerial photography has matured. With aerial photography, whole swaths of forest can be captured within a single shot. Aerial shots, therefore, present themselves as a lucrative, cost-effective solution that reduces the amount of time and labour required. Hence, this study is centred on the use of aerial shots. These aerial shots contain limited multispectral information data, only presenting the crowns of individual trees. Gray Level Cooccurrence Matrix (GLCM) was first applied to the raw data obtained from the drones. With the feature maps obtained from the GLCM across the different channels, the GLCM mean and 5 base channels were identified using Analysis of Variance (ANOVA) to have distinct distributions across classes. Using a modified FaceNet model by Schroff et al., the GLCM mean and 5 base channels were used for FaceNet [1]. FaceNet was used as it was a fully end-to-end, one-shot learning deep learning algorithm that did not require significant amounts of data. The combination of the different methods yielded a test accuracy of 54% across 15 classes.

1 INTRODUCTION

Forest recovery is a topic that has grown in interest in recent years. In view of the increasing rate of deforestation globally, conservatory efforts for forests would greatly benefit from methods of automated monitoring. Aerial pictures taken from Unmanned Aerial Vehicles (UAV) may be used to provide general surveillance of the forests-in-inspection through the use of electro-optic sensors. In addition, electro-optic sensors provide necessary data that contains information from the crowns of trees. Electro-optic information containing the crowns of these trees may then be used to perform tree identification.

Red, green, and blue (RGB) light is captured by these sensors. However, information from reflected RGB light do not present a complete picture in understanding the health of a forest. More bands of light can be captured to supplement the RGB data. These extra bands of light are obtained using multispectral cameras, which capture a variety of wavelengths of light, such as Near Infrared (NIR). In an agricultural context, NIR information can also be used to assess the general health of a forest system.

Tree identification using these multispectral cameras can provide a more comprehensive understanding of the forest. Being able to identify all varieties present in the rainforest is useful in supplementing the general understanding of the health of the biosystem.

In this area of tree identification using tree crowns, Dalponte et al. achieved high efficacies using features extracted by laser scans and hyperspectral information [2]. Dalponte et al. used large datasets across five different trees species. This process is demanding, as much time will have to be spent labelling trees, which is not practicable.

Previous work required the use of classical machine learning algorithms, such as random forests and support vector machines. Any additional classes introduced will require subsequent rounds of training of the models. This process requires more computation. Considering the large number of tree species that exist in the world right now, current solutions may not be feasible moving forward.

Most work have also suggested the incorporation of a wide variety of spectral bands [2] [3]. These spectral bands provide additional features that may be useful in classification. Incorporating a large corpus of spectral bands will require advanced equipment that is costly. Hyperspectral information may therefore be not practical in performing tree identification. Hence, another challenge will be in limiting the number of spectral bands, such as to reduce the overall costs of the project.

2 METHODOLOGY

An end-to-end solution is proposed to circumvent the issues with current solutions. The proposed method will generate signatures of each individual tree species. These characteristic signatures will form a dictionary that will be used for future inferences. Each of these signatures will also be generated using limited amounts of data and on a selection of spectral bands.

2.1 DATA

Multispectral images were captured using a drone. Aerial images were taken at the Chestnut Nature Park using five different spectral bands (Table 1). RGB information was also recorded in varying bandwidths, which can be distinguished into wideband and narrow band. Wideband RGB bands capture a greater bandwidth of RGB channels, while narrow RGB bands capture a smaller band of information.

Band
(Wideband / narrow band) Red
(Wideband / narrow band) Green
(Wideband / narrow band) Blue
Red Edge (RE)
Near Infrared (NIR)

Table 1: Recorded Spectral Bands

The data contains 15 different tree species that has been labelled by the NUS Ecology Department (Table 2). Figure 3 represents the sample of an image using RGB broadbands.

Species
Alstonia Angulstiloba
Calophyllum
Campnosperma Auriculatum
Cinnamomum Iners
Cratoxylum Formosum
Dillenia Suffruticosa
Falcataria Moluccana
Ficus Variegata
Leea Indica
Pennisetum Purpureum
Pommetia Pinnata
Spathodea Campanulatum
Sterculia Parviflora
Syzygium Polyanthum
Terminalia Catappa

Table 2: Tree Species identified at Chestnut Nature Park

Figure 3: Broadband RGB Representation of Chestnut Nature Park

2.2 GRAY LEVEL COOCURRENCE MATRIX

Gray Level Cooccurrence Matrix (GLCM) is a textural analysis method for images. GLCM calculates the relationship between neighbouring pixels using the relative grayscale intensity, distance and angle. Textures generated by GLCM contains useful spatial information which supplements the spectral data provided by the multispectral camera.

Statistical parameters can be obtained from the GLCM. In our study, we obtained the mean (1), variance (2), homogeneity (3), correlation (4) and Angular Second Momentum (ASM) (5). These statistical descriptors may be meaningful and be used in generating individual fingerprints for each of these trees.

$$\mu_k = \sum_{ij} k \cdot c_{ij} \quad - \quad (1)$$

$$\theta_k = \sum_{ij} c_{ij} \cdot (i - \mu_k)^2 \quad - \quad (2)$$

$$\sum_i \sum_j \frac{c_{ij}}{1 + |i - j|} \quad - \quad (3)$$

$$\sum_i \sum_j (i - j)^2 \cdot c_{ij} \quad - \quad (4)$$

$$\sum_i \sum_j c_{ij}^2 \quad - \quad (5)$$

These statistical parameters are applied across all spectral bands indicated in Table 1. The result of these operations result in a total of 40 different feature maps.

2.3 ANALYSIS OF VARIANCE (ANOVA)

ANOVA is a statistical test used to determine the statistical significance between feature maps. Examining the statistical significance allows us to choose better features. Better features better fits models, especially in the context of machine learning models. Performing ANOVA produces the 'F statistic', a ratio comparing the within group variance and the between group variance (6). Significant difference between groups will mean the rejection of the null hypothesis. The null hypothesis is the notion that there is a difference between groups. In this context, feature maps that reject the null hypothesis are features that are statistically significant.

$$F = \frac{Between\ Groups\ Variance}{Within\ Group\ Variance} \quad - \quad (6)$$

To validate statistically significant features, the plots of the feature maps will be plotted. As suggested by the 'F statistic', the highest 'F statistics' are the feature maps that will be used in performing classification.

2.4 PRINCIPAL COMPONENT ANALYSIS

Principal Component Analysis (PCA) is a widely used method in dimension reduction. PCA performs data compression by first determining the principal components of a given system. These principal components are uncorrelated. Most information is also compacted into the first few principal components. Equation (7) details the reassembly of the system by mapping a matrix of data points, *X,* from the original space to a new space using the first few principal components, W_L.

$$T_L = X \cdot W_L \quad - (7)$$

PCA is also useful as a filter. Through the transformation in (7), information that contributes maximally is kept, while redundant information that may be lost is removed.

2.5 FACENET

2.5.1 FACENET WITHOUT MODIFICATIONS

FaceNet is a deep convolutional neural network-based approach developed to perform facial verification and identification [2]. The problems of tree identification and facial recognition are very similar. Both problems face large amounts of classes with limited data present.

FaceNet adopts a unique loss function in training and is the triplet-based loss function developed for Large Margin Nearest Neighbour (LMNN) classification [4]. The triplets are embeddings of faces, reduced into a 128-D vector. During the training phase, a key objective would be to minimise the Euclidean distance between the anchor and positive (8). Both the anchor and positive are of the same class. Concurrently, the next primary goal would be to maximise the Euclidean distance between the anchor and negative (7). The negative and anchor do not share the same identity. Figure 4 illustrates this training phase.

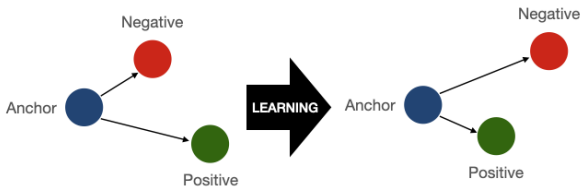

Figure 4: Triplet Loss Minimisation

$$L = \sum_{i}^{N} [\left\| f(x_i^a) - f(x_i^p) \right\|_2^2 - \left\| f(x_i^a) - f(x_i^p) \right\|_2^2 + \propto] \quad - \quad (8)$$

Once the embeddings are generated of the trees, these embeddings are averaged and saved to a dictionary. Each of these embeddings are mapped to a particular class.

2.5.2 MODIFICATIONS MADE TO FACENET

The method of determining true-accepts and false-accepts differ from the original FaceNet model. The original model determines true-accepts by taking a pair of images and comparing the L_2 distance between the pair. A L_2 distance that is lesser at a predefined threshold will mean that the given image is a true-accept, or that the pair of images are similar. Our proposed method differs from this. Instead of using a predefined threshold, the embeddings that have the lowest L_2 norm to the saved dictionary of embeddings, P_{dict}, will mean that image i is associated with the class of trees (9).

$$TA(d) = \{i \in P_{dict}, with\ min(D(i,j))\} \quad \text{- (9)}$$

The last modification made to FaceNet is the use of conversion layers. The original FaceNet was designed to be used for RGB images, constituting of three channels. For the case of tree identification, multiple channels are going to be used. Conversion layers will be added prior to the CNN backbone. Table 4 represents a sample of the use of conversion layers should 10 different feature maps be used.

Conv1 (filter size 3 x 3 with 10 feature maps)
Conv2 (filter size 3 x 3 with 7 feature maps)
Conv3 (filter size 1 x 1 with 3 feature maps)
CNN backbone (InceptionNet)

Table 4: Details of conversion layers

2.5.3 TRAINING OF FACENET

The datasets used in training FaceNet largely consisted of one entry per class. To artificially increase the size of the dataset, each entry is subdivided into four additional entries. Each entry is geometrically divided at the centre into four equal portions.

During the training phase of the FaceNet model, a 3:1 split in training vs validation data was used. Training would stop when training and validation loss converges.

3 RESULTS

3.1 FACENET RESULTS

There are two different sets of data captured over Chestnut Nature Park. One of captured on 20 December 2020, while another was captured on 18 May 2021. Unless otherwise specified, the test results are of May data while training the model on December data.

3.1.1 FIVE BASE CHANNELS

FaceNet was tested against the five base channels. To increase the overall test set size, each test tree was divided into four different trees. Table 5 shows the results for the five base channels when used in FaceNet.

	Precision	Recall	F1-score	Support
Alstonia Angulstiloba	0.55	0.75	0.63	8
Calophyllum	1.00	0.25	0.40	4
Campnosperma Auriculatu	0.00	0.00	0.00	4
Cinnamomum Iners	0.33	0.25	0.29	4
Cratoxylum Formosum	0.31	1.00	0.47	4
Dillenia Suffruticosa	0.50	0.50	0.50	4
Falcataria Moluccana	0.33	0.50	0.40	4
Ficus Variegata	1.00	0.50	0.67	4
Leea Indica	0.00	0.00	0.00	4
Pennisetum Purpureum	1.00	0.75	0.86	4
Pommetia Pinnata	0.43	0.75	0.55	4
Spathodea Campanulatum	0.00	0.00	0.00	4
Sterculia Parviflora	0.83	0.62	0.71	8
Syzygium Polyanthum	0.00	0.00	0.00	4
Terminalia Catappa	0.67	0.50	0.57	4
Accuracy			0.46	68
Macro Average	0.43	0.40	0.38	68
Weighted Average	0.49	0.46	0.43	68

Table 5: Results using five base channels

3.1.2 FIVE BASE CHANNELS AND MEAN FEATURES

Both base and mean features are very well separated. Hence, the performance of FaceNet was evaluated. Table 6 shows the results for the five base channels and the GLCM means when used in FaceNet.

	Precision	Recall	F1-score	Support
Alstonia Angulstiloba	0.40	0.50	0.44	8
Calophyllum	0.80	1.00	0.89	4
Campnosperma Auriculatum	0.00	0.00	0.00	4
Cinnamomum Iners	0.00	0.00	0.00	4
Cratoxylum Formosum	0.22	0.50	0.31	4
Dillenia Suffruticosa	1.00	0.50	0.67	4
Falcataria Moluccana	0.50	0.25	0.33	4
Ficus Variegata	0.00	0.00	0.00	4
Leea Indica	0.80	1.00	0.89	4
Pennisetum Purpureum	1.00	1.00	1.00	4
Pommetia Pinnata	1.00	0.50	0.67	4
Spathodea Campanulatum	0.50	0.75	0.60	4
Sterculia Parviflora	0.64	0.88	0.74	8
Syzygium Polyanthum	1.00	0.25	0.40	4
Terminalia Catappa	0.75	0.75	0.75	4
Accuracy			0.54	68
Macro Average	0.54	0.49	0.48	68
Weighted Average	0.57	0.54	0.52	68

Table 6: Results using five base channels and GLCM means of these channels

3.1.3 ALL GLCM DERIVED FEATURES

All GLCM derived feature maps were used in computing the overall accuracy of FaceNet. Table 7 details the results.

	Precision	Recall	F1-score	Support
Alstonia Angulstiloba	0.33	0.12	0.18	8
Calophyllum	1.00	1.00	1.00	4
Campnosperma Auriculatum	0.00	0.00	0.00	4
Cinnamomum Iners	0.00	0.00	0.00	4
Cratoxylum Formosum	0.11	0.25	0.15	4
Dillenia Suffruticosa	0.20	0.50	0.29	4
Falcataria Moluccana	0.00	0.00	0.00	4
Ficus Variegata	0.00	0.00	0.00	4
Leea Indica	0.00	0.00	0.00	4
Pennisetum Purpureum	0.00	0.00	0.00	4
Pommetia Pinnata	0.00	0.00	0.00	4
Spathodea Campanulatum	0.14	0.50	0.22	4
Sterculia Parviflora	0.25	0.12	0.17	8
Syzygium Polyanthum	0.00	0.00	0.00	4
Terminalia Catappa	0.67	0.50	0.57	4
Accuracy			0.19	68
Macro Average	0.17	0.19	0.16	68
Weighted Average	0.19	0.19	0.17	68

Table 7: Results using five base channels and all GLCM derived features

3.1.4 TEST OF ROBUSTNESS

To validate the robustness of the devised model, this set was trained on May data, while tested on December data. For comparison, the best performing combination of feature maps was used – using the base channels and GLCM mean features.

	Precision	Recall	F1-score	Support
Alstonia Angulstiloba	0.60	0.75	0.67	4
Calophyllum	1.00	0.25	0.40	4
Campnosperma Auriculatum	0.00	0.00	0.00	4
Cinnamomum Iners	0.14	0.25	0.18	4
Cratoxylum Formosum	1.00	0.25	0.40	4
Dillenia Suffruticosa	0.00	0.00	0.00	4
Falcataria Moluccana	1.00	0.25	0.40	4
Ficus Variegata	0.40	0.50	0.44	4
Leea Indica	1.00	0.50	0.67	4
Pennisetum Purpureum	0.50	0.75	0.60	4
Pommetia Pinnata	0.14	0.25	0.18	4
Spathodea Campanulatum	0.33	0.50	0.40	4
Sterculia Parviflora	0.38	0.75	0.50	8
Syzygium Polyanthum	0.50	0.25	0.33	4
Terminalia Catappa	1.00	0.25	0.40	4
Accuracy			0.39	64
Macro Average	0.50	0.34	0.35	64
Weighted Average	0.52	0.39	0.38	64

Table 8: Five base channels and mean features, training on May data and testing on December data

3.1.5 EXPANSION OF DATASET

The dataset was expanded to include 19 classes. Subsequent generations of GLCM were also generated. This was done by modifying parameters such as the size and orientation of the window. Lastly, the base channels and GLCM means were used as the combination for the input. For this dataset, the GLCMs were generated with the following parameters:

- Radius: 7
- Step Size: 15
- Bin Size: 128

3.1.5.1 FILTERING USING PCA TO IMPROVE SEPARATION OF FEATURE MAPS

Another step applied to the expanded dataset was the application of PCA to function as a filter. The rationale for a filter in this case was to eliminate noisy data that may affect the separation within feature maps. Hence, PCA was applied directly onto each feature map. After PCA was performed, the first five principal components were reconstructed to the original size of each feature map. In this case, the inputs to FaceNet will remain the same, and the structure of FaceNet would not change. With better separation between different feature maps, the classification result is expected to be better. This was observed separately using feature maps identified using GLCM of the following parameters:

- Radius: 3
- Step Size: 7
- Bin Size: 128

Prior to the application of PCA for GLCMs using the above parameters, the accuracy between 19 classes yielded an accuracy of 27%. However, applying PCA and keeping only the first five principal components resulted in an increase in accuracy to 33%.

3.1.5.2 TEST RESULTS FROM EXPANSION

For the test data, a support of 4 was maintained throughout all species of trees. As all the tree species do not have more than 4 samples each in the test set, tree species that have more than one crown samples will have one or two of the quadrants selected. Table 9 explains in greater detail the allocation of the four supports used in the test data.

Number of Instances per Species	Allocation Details
1	All quadrants used in test data
2	Two quadrants from one instance, another two quadrants from the other instance
3	Two quadrants from one instance, one quadrant from another instance, and the last quadrant from the last unused instance
4	One quadrant from each of the four instances

Table 9: Allocation of testing data to achieve a support of four

	Precision	Recall	F1-score	Support
Alstonia Angulstiloba	0.50	0.50	0.50	4
Bridelia Sp	1.00	0.50	0.67	4
Calophyllum	0.40	1.00	0.57	4
Campnosperma Auriculatum	0.00	0.00	0.00	4
Cinnamomum Iners	0.25	0.25	0.25	4
Claoxylon Indicum	0.00	0.00	0.00	4
Clausena Excavata	0.17	0.25	0.20	4
Dillenia Suffruticosa	0.50	0.25	0.33	4
Falcataria Moluccana	0.40	0.50	0.44	4
Ficus Variegata	0.33	0.50	0.40	4
Leea Indica	0.75	0.75	0.75	4
Pennisetum Purpureum	0.00	0.00	0.00	4
Pometia Pinnata	0.00	0.00	0.00	4
Sandoricum Koetjape	0.33	0.25	0.29	4
Shorea Leprosula	0.50	0.25	0.33	4
Spathodea Campanulatum	0.00	0.00	0.00	4
Sterculia Parviflora	0.25	0.25	0.25	4
Syzygium Polyanthum	0.40	1.00	0.57	4
Terminalia Catappa	0.80	1.00	0.89	4
Accuracy			0.38	76
Macro Average	0.35	0.38	0.34	76
Weighted Average	0.35	0.38	0.34	76

Table 10: Five base channels and GLCM mean features on the expanded dataset

3.1.5.3 TESTING ROBUSTNESS OF DATA FROM EXPANDING DATASET

In a similar manner to the procedures in section 3.1.4, the robustness of the devised model was tested on the expanded dataset. This set was trained on May data, and subsequently tested on December data. The base channels and GLCM mean features were used as inputs to the FaceNet.

	Precision	Recall	F1-score	Support
Alstonia Angulstiloba	0.50	0.50	0.50	4
Bridelia Sp	0.00	0.00	0.00	4
Calophyllum	1.00	1.00	1.00	4
Campnosperma Auriculatum	0.50	0.25	0.33	4
Cinnamomum Iners	0.00	0.00	0.00	4
Claoxylon Indicum	0.50	0.25	0.33	4
Clausena Excavata	0.25	0.25	0.25	4
Dillenia Suffruticosa	0.50	0.25	0.33	4
Falcataria Moluccana	0.00	0.00	0.00	4
Ficus Variegata	0.00	0.00	0.00	4
Leea Indica	0.50	0.50	0.50	4
Pennisetum Purpureum	1.00	0.50	0.67	4
Pometia Pinnata	0.00	0.00	0.00	4
Sandoricum Koetjape	0.50	0.50	0.50	4
Shorea Leprosula	1.00	0.25	0.40	4

Spathodea Campanulatum	0.22	0.50	0.31	4
Sterculia Parviflora	0.20	0.25	0.22	4
Syzygium Polyanthum	0.25	0.50	0.33	4
Terminalia Catappa	0.33	0.25	0.29	4
Accuracy			0.30	76
Macro Average	0.38	0.30	0.31	76
Weighted Average	0.38	0.30	0.31	76

Table 11: Five base channels and GLCM mean features on the expanded dataset, training on May data and testing on December data

4 DISCUSSION

Given the limitations imposed by the problem, the modified FaceNet model has achieved a satisfactory level of results. FaceNet was able to discern some species with relative low levels of confusion, while some species with higher levels of confusion. Species such as *Campnosperma Auriculatum* and *Cinnamomum Iners* have poor accuracies, while *Pennisetum Purpureum* and *Calophyllum* have nearly perfect identification. It highlights that FaceNet is selective in performing tree identification, such that only certain species are generalised well. As a result, the result as reflected in Table 6 averages out to be of 54% accuracy using the feature maps produced by the five base channels and the GLCM means of each of the five channels.

This trend could be explained by the high levels of confusion between species, especially given the large number of classes introduced for this study. More features may need to be introduced, such that there is a greater degree of separation between classes. The current constraints of five spectral bands maybe insufficient to create this level of separation. Additional texture maps, on top of the current GLCM features, may need to be generated to compliment the lack of distinction between classes.

Reducing the total number of feature maps used in FaceNet also results in a decrease in accuracy. In Table 5, using only the GLCM of the five base channels reflect a drop in accuracy to 46%. This accentuates the need for more and statistically significant feature maps to help FaceNet to perform tree identification.

4.1 DATA INCONSISTENCIES AND MODEL GENERALISATION

The results of Table 6 and 8 shows that there are differences between the May and December data. Training on December data and testing on May data brought about recall and precision greater than 50% in 8 / 15 of tree species. Training on May and testing on December data, however, resulted in a drop in the recall and precision, where 3 / 15 tree species maintained a recall and precision greater than 50%. Accuracies also fell from 54% to 39% when the training dataset was on the May dataset while the testing dataset was on the December dataset. The results of Table 10 and 11 further back this observation, with accuracy reducing from 38% to 30% when the training dataset was on May dataset and tests were made on the December dataset.

These results do strongly indicate data inconsistencies across December and May datasets. These differences in data could be due to factors such as lighting conditions and the health of the trees. It also suggests that there is the inability for the model to generalise and classify the correct species of trees.

4.2 IMPORTANCE OF FEATURE SELECTION

Peak performance was obtained when using the feature maps of the GLCM and its corresponding mean for each of the five channels. This is consistent with the results of ANOVA, which indicated that these ten feature maps were very well separated.

As shown in Table 7, using all GLCM derived features caused a significant dip in accuracies. The accuracy of the model dropped to 19%, with high levels of confusion across many tree species. This may be explained by the lack of separation between each of the feature maps, which introduces a higher level of confusion to the model.

4.3 TRAINING UNCERTAINTIES

Another limitation arising from the lack of data, is the inability of the model to perform conclusive validation steps during training. A 3:1 split in training and validation data indicates that the validation step may consist of a single instance per class. As a result, the validation step may not be sufficient in understanding on how well the model is fitted to the training data. The current results presented may be tested on trained models that have not fitted well to the data. As a result, the model might not have arrived to its most optimal fit, and could be generalised further to produce better results.

The current conversion layers used to compress multiple feature maps into smaller dimensions may also not be effectively in reducing the total number of dimensions. An observation is that in training models that are compressing greater number of feature maps require more amounts of training, as training and validation losses appear not to be reducing much after subsequent rounds of training.

5 CONCLUSION

The modified FaceNet model may not be the most robust solution for the problem. It is apparent from the results that the constraints of the problem negatively affect the performance of the model. Across data with varying conditions, the proposed solution is not able to generalise well with the few single instances of classes provided. With single instances of trees available for training, validating the classes may not be enough in having a complete understanding of the fit of the model. These problems are centred on the lack of data. However, there are still other areas to explore to boost the reliability and efficiency of the model.

1) Increase the total number of statistically significant feature maps. This involves finding more ways of generating feature maps that are statistically significant. One way would be to explore other GLCM statistical parameters. Another way would be to explore other forms of textural feature extractions, such as image transforms to study the spatial frequency spectrum. This form of transforms may supplement the generated GLCMs and aid in the classification of a greater variety of tree species.

2) Incrementally increase the amount of data for training. Having a single instance for each class has shown that it is likely insufficient to produce a reliable model. We need a clearer idea of the general fit of the model. Few more instances of the same class not only allow more training data for the model, but also more data for validating the fit of the model.

3) Use data compression neural networks to compress the large amounts of feature maps into smaller dimensions. As observed, the few layers used in compressing the total number of feature maps may not be effective. Autoencoders and Principal Component Analysis (PCA) have been shown as possible methods in projecting the multiple feature maps into smaller dimensions [5]. A better compression may allow the model to fit better with more defined features between classes.

These suggestions to the training phase are not exhaustive. While improvements may be made to yield better results, they are not to suggest that the modified FaceNet does not work. The results from the modified FaceNet model are promising and is able to predict with near perfection for some classes based on single instances of trees.

ACKNOWLEDGMENT

I would like to acknowledge the funding support from Nanyang Technological University – URECA Undergraduate Research Programme for this research project.

I would like to also acknowledge the support from NUS Ecology Department in providing labels for the trees.

Dr. Ji-jon Sit, the supervisor for this URECA project, who has been so diligent in helping the team out throughout to this project.

Lastly, I would like to thank the Undergraduate Research Assistants, John Chang and Benjamin Chew, for their previous contributions to the project.

REFERENCES

[1] F. Schroff, D. Kalenichenko and J. Philbin, "FaceNet: A unified embedding for face recognition and clustering", 2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2015. Available: 10.1109/cvpr.2015.7298682 [Accessed 21 June 2022].

[2] M. Dalponte, L. Frizzera and D. Gianelle, "Individual tree crown delineation and tree species classification with hyperspectral and LiDAR data", *PeerJ*, vol. 6, p. e6227, 2019. Available: 10.7717/peerj.6227 [Accessed 21 June 2022].

[3] A. Dietmaier, G. McDermid, M. Rahman, J. Linke and R. Ludwig, "Comparison of LiDAR and Digital Aerial Photogrammetry for Characterizing Canopy Openings in the Boreal Forest of Northern Alberta", *Remote Sensing*, vol. 11, no. 16, p. 1919, 2019. Available: 10.3390/rs11161919 [Accessed 21 June 2022].

[4] K. Weinberger and L. Saul, "Distance Metric Learning for Large Margin Nearest Neighbor Classification", *Journal of Machine Learning Research*, vol. 10, no. 2009, pp. 207-244, 2009. [Accessed 21 June 2022].

[5] Q. Fournier and D. Aloise, "Empirical Comparison between Autoencoders and Traditional Dimensionality Reduction Methods", *2019 IEEE Second International Conference on Artificial Intelligence and Knowledge Engineering (AIKE)*, 2019. Available: 10.1109/aike.2019.00044 [Accessed 21 June 2022].