

# Mangrove Classification using Semi-Supervised Learning

*Ashlesha Vaidya, Sidharth Suresh*

## 1. Abstract

Mangrove forests reduce global greenhouse emissions by 12%, these rich ecosystems support the planet and the people in many unique ways but are declining at a rapid rate due to deforestation. To monitor and track mangroves, geospatial data captured by drones is used. Labelling these for machine learning purposes is a tedious task and hence in order to use large amounts of labelled data for segmentation of geospatial data, semi supervised machine learning methods are employed. A model with label propagation technique alongside a supervised classifier is built to classify mangrove and non-mangrove pixels in aerial images. The accuracy of the different models ranges from 74% to 89% depending on various factors like the amount of labelled data. Overall, the study shows that even though a semi-supervised model has lesser accuracy than previously implemented CNNs but they largely automate the process of labelling the aerial images in an efficient manner.

## 2. Introduction

### 2.1 Background

Mangroves are trees and shrubs that have adapted to grow in the intertidal zone along subtropical coastlines. Mangroves protect shorelines from damaging storm and hurricane winds, waves, and floods. They reduce the cost of coastal defense by forming an outer layer protection[1]. Mangroves also help prevent erosion by stabilizing sediments with their tangled root systems. They maintain water quality and clarity, filtering pollutants and trapping sediments originating from land. Mangroves are among the most carbon-rich biomes, facilitating accumulation of fine particles and fostering rapid rates of sediment accretion[2]. they sequester 2 times more carbon compared to tropical rainforests[3].

In order to be protected, mangrove regions need to be identified first. Hence, these mangrove regions are segmented on geospatial images. This will help conservationists and researchers for mangroves conservation. The main idea of this project is to identify mangrove regions from drone images. Fig.1 shows an example of drone image.



Fig.1 - Captured drone image of a mangrove

### 2.2 Previous Work

As shown in fig.2, high resolution drone images are captured and mangrove regions in these images need to be identified. But manually identifying these regions is time consuming as assigning labels to an image of size  $\sim(15000 \times 15000)$  pixels is a tedious task.

This problem has been tackled previously by implementing various classification models such as Object[6] and Pixel[5] based Support Vector Machines, Convolutional Neural Network for segmentation[7] of mangrove regions[8]. The CNN model developed by the Mangrove team UCSD has developed a InceptionResnetV2[4] CNN model that has an accuracy of 97-99%. Fig 2 shows the results obtained from each of the classifier models and their corresponding comparison to the ground truth values.

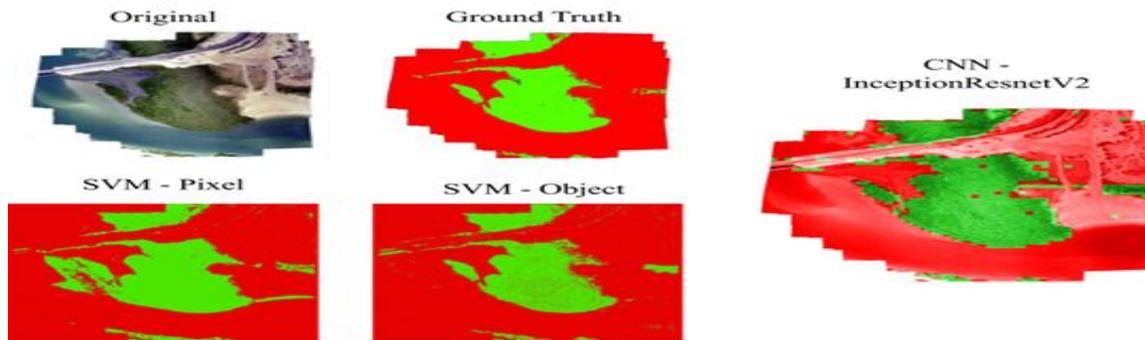


Fig.2 - Existing Classification models and their result

In these methods, only the labelled data is used and unlabelled data are not used for any purpose. The training data for these models had to be hand labelled manually and it requires a large amount of time for the same. This process involves dividing each image into tiles using a polygon file, creation of label raster and manually assigning boundaries to mangrove regions.

Thus, the project aims to make use of these unlabelled data along with the limited labelled data available for the classification of mangroves.

### 2.3 Our approach

Semi-Supervised Learning is a Learning approach that makes use of less number of labelled data and a large number of unlabelled data for training. Many of the previous works which implements SSL techniques make few underlying assumptions about the dataset - smoothness, low-density[10] and cluster assumptions[16]. Smoothness assumption is that if 2 samples are close in the input space then their labels should be the same. The low-density assumption states that the decision boundary should lie in an area where few data points are observed. These assumptions are made to ensure that the decision boundary can separate different classes of data.

Semi-Supervised Learning has already been proven to be a powerful paradigm for leveraging unlabeled data to mitigate the reliance on large labelled datasets[9] in Image classification techniques. This project aims to implement a Semi-Supervised model for image segmentation of drone images of mangrove sites.

Semantic segmentation using Semi-Supervised Learning and Neural Networks have been previously done on the ADAC dataset[11] using the process of co-training where models are trained on a subset of the annotated data and the use the non-annotated data to exchange information with each other. The use of Semi-Supervised Learning and pseudo labels for Image Classification on hyperspectral images[12] and aerial images[13] already exist, we aim to implement a Semi-Supervised Machine Learning Model using Label Propagation for semantic segmentation of drone images.

This project mainly covers 3 main steps involved in Mangrove Classification using Semi-Supervised Learning.

1. Data Preparation and Data Pre-processing - This step involves the dividing of a large sized drone image into tiles and extracting the ground truth values and generating labels.
2. Label Propagation - This step aims to label the unlabelled data so that they can be combined with the labelled data to train a classifier. Label Propagation techniques such as Clustering, Pseudo Labelling and Graph based methods have been tested and their implementation and results are discussed in the Technical Material Section.
3. Classifier - This is the final part which completes the Semi-Supervised Model. The labelled data along with the data labelled through label propagation is used to train a supervised classifier. Data Points are chosen at random and are used to train the classifier. Different classifiers like random forests, SVM, MLP are experimented with and talked about in the later sections of the report.

We have also improved upon the given model by making use of different filters as feature values as inputs to the supervised classifier. This increases the complexity of the model and gives it more information about any particular pixel in a geospatial image by capturing properties of its neighbors. The performance of the model is evaluated by accuracy metrics for different amounts of labelled data. This shows how the model obviously improves when given more labelled training data. Ultimately, we focussed on improving the model for getting better results when the amount of labelled data is less. This we believe can be a very useful tool as it will save a lot of time and labour spend on manually labelling the geospatial images of mangrove ecosystems.

## 3. Technical Material

This section of the paper delineates the technical part of the project. We have built a semi-supervised learning model for mangrove classification.

### 3.1 Data Collection

Our data consists of geospatial images of mangrove ecosystems captured by a DJI Phantom 4 Pro unmanned aerial vehicle (UAV) with its on-board camera. These images were captured at 4K Resolution (3840 x 2160 pixels) at heights of both 10m and 120m above the highest point of the mangrove canopy. These aerial images were already orthorectified and labelled by volunteers to train our machine learning models. These images were converted into orthomosaics by the mangrove monitoring team. They did this using Agisoft Photoscan 1.4.2, which is a photogrammetric software that can be used to generate images for GIS or other imaging applications. These orthomosaics are an accurate representation of an area in our case the mangrove ecosystem. This is created by stitching many images together and orthorectifying them. The fig.3 shows a drone captured image and fig.4 shows an example of an orthomosaic obtained after the processing steps of orthorectification.



Fig.3 - Drone captured image

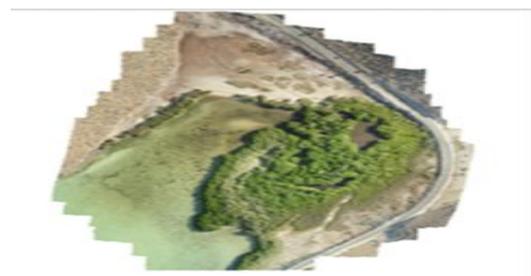


Fig.4 - Orthomosaic of La Paz

We have considered the La Paz, Baja California Sur site(Fig.4) for our training and testing purposes. This site was chosen because it has a lot of variability and brings out few challenges as the model suffers due to blocky areas of water which look much like the mangroves from an aerial view.

The labelling of these images is quite tedious and takes a lot of manpower. Fig.5a and 5b show an example of unlabelled and labelled orthomosaic. The unlabelled orthomosaic is labelled in QGIS with polygons marking the boundaries of the mangroves. The labels are stored as shapefiles. QGIS is also used to fix geometries of the shapefiles before it can be used as data for our semi-supervised model.



Fig 5a. - Unlabelled orthomosaic

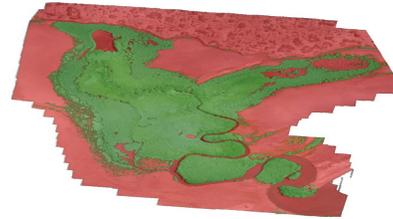


Fig 5b - Labelled orthomosaic

### 3.2 Data Pre-processing

As mentioned above, QGIS is used to fix geometries of polygon files before we could use it for any of the preprocessing steps. After this we perform the following pre-processing steps.

#### 3.2.1 Retiling

Each orthomosaic is pretty huge and has data in the range of 1-4GB. Using these images by themselves for training any machine learning model is impractical. Thus these bigger orthomosaics need to be segmented. However segmenting them is itself a difficult problem[14]. Another solution for this is to break down into smaller tiles of 4096x4096 and then perform coarse classification of the image by classifying the tiles instead of pixels as was introduced in [15]. The larger image files are broken down into tiles using a polygon file for the orthomosaic under consideration. Fig.6a and Fig.6b show how the broken tiles look like.



Fig.6a - Original orthomosaic



Fig.6b - Orthomosaic broken into tiles

### 3.2.2 Generating label rasters using QGIS

We have labelled shapefiles (shown in Fig.7) but since we can't directly use shapefiles in our model we converted these to tiff image files using QGIS. Each tile will have its own labelled tiff file. The label raster is vectorized and an intersection is found with the polygon of each individual tile. This gives a label raster for each tile which is then rasterized. Fig.7 shows the label rasters for the tiles shown in Fig.6b. All black colored pixels represent a mangrove pixel and all non-mangrove pixels are represented by the white pixels. Now we have rasters for tiles and their respective labels.

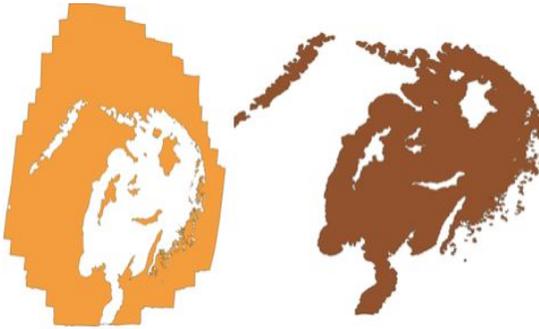


Fig 7. Shapefiles for labels



Fig 8. : Labelled raster

### 3.2.3 Reading in the tiles

The broken down tiles are using opencv and GDAL. These are read in as numpy arrays. Each tile has 4 bands- RGB and an alpha band to represent transparency. An alpha value of 255 denotes no-data and all such no-data pixels are removed before feeding it to our model.

### 3.2.4 Generating labelled image matrices from the masks

We have a labelled raster which acts as a mask and labels need to be extracted from these masks. In order to do this, the labelled raster file is read in and all pixels with RGB value (0,0,0) are marked as 1 while for (255,255,255) are marked as 0. We thus have 0/1 arrays representing mangrove and non-mangrove pixels ready to be used in our machine learning model.

The training data is thus tiles of 4096x4096 with 3 features - RGB and the target is a tile of 4096x4096 with values either 0 or 1. Also, since we are working with La Paz data and this is a fully labelled site, we replace random pixels with -1 to denote an unavailable data point. The number of unlabelled points is chosen such that the number of samples available for both the classes (mangroves and non-mangroves) are equal. The number of unlabelled samples can be varied and the model is experimented with different numbers of unlabelled data points. The performance has been shown in the later sections.

## 3.3 Cluster-then-label approach

This is one of our preliminary approaches we tried for the semi supervised classification. A cluster-then-label method as proposed by [16] would be potentially used to identify high-density regions in the UAV imagery which would then be used to help a supervised SVM in finding the decision boundary and classify pixels into mangroves and non-mangroves. Before we implemented this approach we looked at our data distribution using unsupervised clustering. We fed our tiles to a clustering algorithm and visualized how well the model performed and clustered the mangrove and non-mangrove pixels.

We tried a couple of clustering algorithms like batched k-means and gaussian mixture models for clustering the pixels in the aerial imagery. The mini-batch k-means algorithm of Sculley (2010), works as follows. Centroids are initialised as a random selection of k samples. Then at every iteration, a small number of samples are selected uniformly at random and assigned to clusters. Cluster centroids are updated as the mean of all samples ever assigned to them[17]. The second clustering algorithm - Gaussian mixture model (GMM) finds a Gaussian probability distribution that best model any input dataset. This also accounts for variance in the distribution of data. But as seen in Fig.9 and Fig.10 we can see that the unsupervised clustering by itself does not give satisfactory results. This can probably be due to the pixel properties of the mangroves being similar to the pixels properties of blocky areas of water.

We also tried another clustering algorithm (DBSCAN) which we thought would have performed better as it is density based clustering. In the drone captured images, we can see that the mangroves are densely clustered together, but since this clustering method is inefficient for large datasets due to it's recursive nature, we had speed issues implementing this. We were not able to successfully implement this.

Labelling these clusters as mangrove and non mangroves and feeding this to a supervised classifier will not give good results as there will be a lot of errors in the labelling due to incorrect formation.

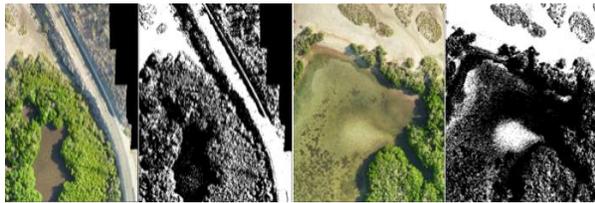


Fig.9 - Mini-batch k means clustering results

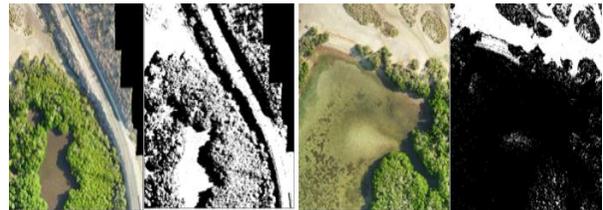


Fig.10 - Gaussian Mixture Model Clustering results

### 3.4 Model Architecture

Since clustering doesn't work well by itself, we decided to follow a different approach. The Fig.11 shows our overall model architecture.

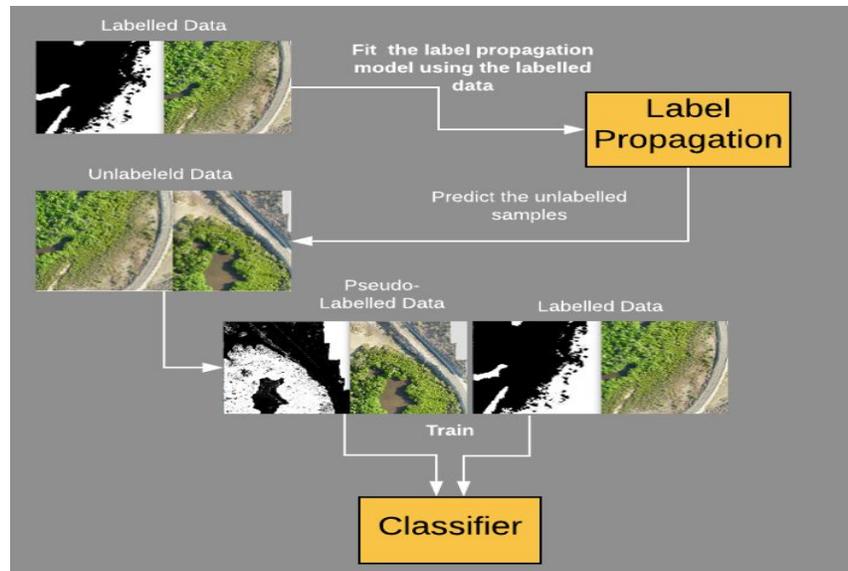


Fig. 11 Model Architecture

The workflow of the model can be explained in the following steps:

**STEP 1:** Separate the labelled data points from the available data points. The labelled samples are fed into a label propagation model which is fit on these samples.

Label propagation is a semi automatic annotation process for labeling a large amount of unlabeled data. The objective of label propagation algorithms is to spread the labels from a small set of labeled data to a larger set of unlabeled data[19] This can be done in multiple ways. We experimented with 2 ways to do this- pseudo labelling and graph based label propagation. The second is described as an improvement on this model.

### 3.4.1 Pseudo-Labeling as a label propagation method

This label propagation method is nothing but a supervised classifier which is trained on the available labelled data. Pseudo-labelling is a simple and efficient semi-supervised technique[18]. Few classifiers are experimented with, to be used as a classifier for pseudo-labelling.

**STEP 2:** The trained label propagation model is used to predict the labels for all the unlabeled samples.  
**STEP 3:** We now have image pixels some of which were originally labelled and some are pseudo labelled with the above mentioned label propagator. The most confident predictions of this model are chosen and the supervised classifier is retrained with these pseudo labelled points alongside the originally available labelled data.  
**STEP 4:** This final retrained classifier can be used to classify new pixels into mangroves or non mangrove classes.

### 3.4.2 Classifier

The classifier used for label spreading and the final pseudo labelling classifier is a random forest classifier. We first tried to implement Support vector machines as our classifier but due to the complexity of the model it crashes when trained on our large dataset. Hence we decided to switch to a more efficient model - random forest. Random forest consists of a large number of individual decision trees that operate as an ensemble. Each individual tree in the random forest spits out a class prediction and the class with the most votes becomes our model's prediction[21]. Different classifiers were implemented like Random Forests, SVM, MLP neural networks , Adaboost classifiers.From all these the ensemble based random forest performed the best in terms of accuracy and efficiency. Thus, random forests were chosen to perform different evaluations.

## 3.5 Evaluating Performance

The semi supervised pseudo-labelling classifier is tested on multiple test samples taken from the same site. Ther performance of the model is tested on different ratios of labelled:total data set size .The performance is observed as the labelled data increases. The Fig.12 shows how the random forest based pseudo labelling semi-supervised model(RF-PL SSL) performs on different amounts of labelled data.

Classifier	10% labelled tiles	25 % labelled data	50% Labelled tiles	70% labelled tiles
RF-PL SSL	73.44%	75%	83.42%	86.38%

Fig 12: Performance of RF-PL SSL on different amounts of labelled data

The Fig.13a and Fig.13b shows the visualization of the classification on all the tiles of the training site when 50% of the labelled data was provided to the model. The white pixels in 11b shows the mangrove classified pixels. It can be observed that the model suffers due to the presence of the blocky areas of water in the aerial images.

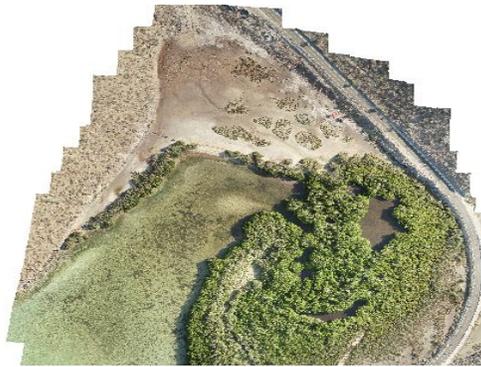


Fig.13a - Original geospatial image

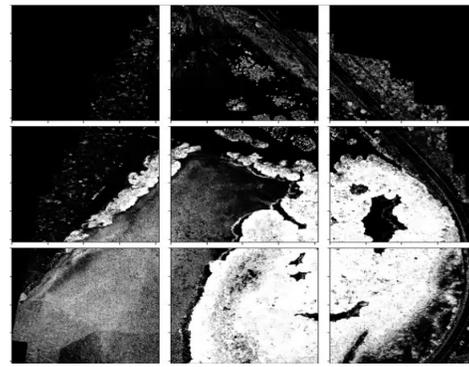


Fig.13b - Classified by RF-PL SSL

### 3.6 Improvements

In order to improve the performance of the current model we tried a couple of things. Since the current model only has 3 features, it seems that increasing the complexity of the model by increasing the number of features may be beneficial. We also tried using a more complex label propagation scheme to see if we can get an improvement on the current model.

#### 3.6.1 Feature Engineering

We add features to the current input set to increase the complexity of the model. This is done by using some filter values present in scikit learn as added features. The used filters include roberts, sobel, scharr, prewitt, gabor. These filters see how the value of a certain feature is affected by its surrounding values. The Fig.14 shows the performance of RF-PL SSL with filters. An improvement in the performance is observed.

Classifier	10% labelled tiles	25 % labelled data	50% Labelled tiles	70% labelled tiles
RF-PL SSL with filters	75.87%	78.82%	85.91%	88.87%

Fig.14 - Performance of RF-PL SSL with filters on different amounts of labelled data

#### 3.6.2 Graph based label propagation

In a graph based label propagation a graph is constructed, often from independent data, based on similarity. As such, edges tend to connect instances with the same class label[20]. Label propagation models in scikit learn have two built-in kernel methods- knn and rbf. Choice of kernel affects both scalability and performance of the algorithms. The rbf kernel forms a dense matrix while knn forms a sparse matrix. Thus our experimentations have been performed using the knn kernel because it is more memory friendly.

When graph label propagation is used along with Random forest classifier, we get an improvement in the accuracy of around 1-2% than our previous model with filters.

## 4. Milestones

P1 - High priority, P2 - Normal priority, P3- Low Priority

Timeline	Task	Priority	Handled by
Week 3, Week 4	Getting started <ul style="list-style-type: none"> <li>• Determining project approach</li> <li>• Literature review - Semi Supervised Learning GMMs</li> <li>• Set up QGIS for visualizing images</li> <li>• Set up GDAL and Azure VMs</li> </ul>	P1	Ashlesha, Sidharth
Week 5	Clustering of the tiles <ul style="list-style-type: none"> <li>• Dividing image into tiles</li> <li>• Clustering of the tiles</li> <li>• Experimenting with different clustering algorithms</li> </ul>	P1	Ashlesha, Sidharth
Week 6	Classification <ul style="list-style-type: none"> <li>• Use labelled tiles available to implement SVM classification algorithm (Supervised Learning)</li> </ul>	P1	Ashlesha, Sidharth
Week 7	Classification with the clusters <ul style="list-style-type: none"> <li>• Incorporate the clustered tiles into the dataset and train the classification model (Semi-Supervised Learning)</li> </ul>	P1	Ashlesha, Sidharth
Week 8	Performance Evaluation <ul style="list-style-type: none"> <li>• Visualization of the model's performance on test data</li> <li>• Evaluate the accuracy</li> <li>• Compare the performance with already existing CNN model</li> </ul>	P2	Ashlesha, Sidharth
Week 9	More testing <ul style="list-style-type: none"> <li>• Look for further scope of improvement to increase performance</li> <li>• Test the model on new data (Different site)</li> </ul>	P3	Ashlesha, Sidharth
Week 10	Making the video/final documentation	P1	Ashlesha, Sidharth

Fig.15 - Initial Milestone Table

The above table (Fig.15) contains the initial plan and milestones set for the quarter. The following were the changes made as the quarter progressed -

### 4.1 A new task- Label Propagation and Implementation of Random Forest Classifier

Week 7	Classification with the clusters <ul style="list-style-type: none"> <li>• Label Propagation</li> <li>• Implement Random Forest Classifier</li> <li>• Incorporate the labelled tiles into the dataset and train the classification model (Semi-Supervised Learning)</li> </ul>	P1	Ashlesha, Sidharth
--------	---	----	-----------------------

The milestones and deliverables planned Week 3-6 were achieved successfully. The data preprocessing was completed, different clustering methods were tested out and results were obtained, the SVM classifier was ready at the end of week 6.

We set out to implement our “Cluster then Classify Approach”, but this hit a dead end when the pixel wise clustering techniques did not give accurate results in spite of various methods tried. Hence we had to look for alternatives such as new label propagation techniques such as pseudo labelling and graph based label propagation methods.

Hence a new task, Label Propagation was added to look for better alternatives to assign labels to unlabelled data more accurately.

Another task added was the implementation of Random Forest Classifier, the Random Forest Classifier used decision trees and yielded better results at a lesser runtime compared to the SVM Classifier, when both the models were tested on a random set of images. Also SVM crashes for larger training sets due to the complexity of the model. Hence, we chose to go ahead with the Random Forest Classifier and week 7 was dedicated to implement the Classifier and enable it to run on our dataset and then integrate the entire classifier along with the label propagation model.

#### 4.2 Scope of Improvement carried over to week9 and Label Propagation to Week8

Week 8	Performance Evaluation <ul style="list-style-type: none"> <li>● Label Propagation</li> <li>● Evaluate the accuracy</li> <li>● Look for further scope of improvement to increase performance (Feature extraction)</li> <li>● Visualization of the model’s performance on test data</li> </ul>	P2	Ashlesha, Sidharth
Week 9	More testing <ul style="list-style-type: none"> <li>● Look for further scope of improvement to increase performance</li> <li>● Test the model on new data (Different site)</li> </ul>	P2	Ashlesha, Sidharth

Due to the addition of a new task, the milestones set for Week 7 and Week 8 had to be carried over to Week 9. While different label propagation techniques were tried and tested, the Semi-Supervised Model with each of the label propagation techniques were also parallelly built to measure the accuracy of the entire model.

Simultaneously, different parameters of the classifier were modified and tweaked to improve accuracy. Different feature extraction techniques were also implemented and the model was modified to obtain better results and the whole process was carried over to the following week.

#### 4.3 Testing the model on a different Site

One of the milestones set for Week 9 was to test the model on a new dataset, this milestone was not completed as improving the current working model’s performance and to find a better label propagation technique was of higher priority. The main goal set this was to implement a robust Semi-Supervised Model which was of higher priority.

Testing on a new site also meant we had to generate image tiles and shape files to extract the labels of the new site. This is a time consuming task. Hence, this milestone was not completed as it was of low priority and it would be time consuming.

But from our knowledge of the dataset of different mangrove sites and from studying previous literature, we expect the Semi Supervised Model trained on a particular site to provide similar or less accurate results depending on how varied the vegetation and the pixel properties are.

Something that we would like to look more into will be the behaviour of the model when data from two different sites are mixed and are used for training. This would give us insights about whether the model can be used for classification of different sites together or if a separate model needs to be trained from scratch for a different site.

We planned to achieve the following deliverables in the given timelines -

- Week 5 : github repo with the tiling code, tiles of image
- Week 7: clustering code with results, classification
- Week 9 : semi supervised model

We have successfully delivered these within the given deadlines.

All the milestones set were worked on together since each milestone has a lot of sub parts to be completed. Previously, our Minimum Viable Product was a Semi Supervised “Cluster then Classify” model trained on our dataset. The input to this model is the actual drone captured image divided into tile along with the labelled tile images generated by clustering. But then once we knew we obtained better results from the Random Forest Classifier along with the Pseudo Labeling, we decided to set the Pseudo Labelling Random Forest Classifier model as our MVP and build on this model. This was done to raise the standards set and to enable us to achieve better results.

## 5. Conclusion

In conclusion, we were able to successfully build a tool for labelling the mangrove pixels in geospatial images. For this we used a label propagation model alongside a supervised classifier to classify pixels based on the available labelled training data. In order to feed our images to the model they were pre-processed and a numpy array representing the images was created for both the actual images and the labels. After experimenting with different classifiers and adding some improvements like adding features to the existing training set, the accuracy obtained on the entire site comes to be around 82.3% with a deviation of 5% depending of the amount of labelled data(this ranged from 10%-70% in our tests).There is a lot of scope of improvement to this project. In the future we can look at how the model performs on neural network classifiers and also test the model on a completely different site. The properties of the pixels vary from site to site and hence it will be an interesting study to see how well our model generalizes.

## 6. References

1. H.J. Verhagen and Tran thi Loi, “The use of mangroves in coastal protection”, 8th International Conference on Coastal and Port Engineering in Developing Countries, 2012
2. Daniel M Alongi, “Carbon Sequestration in Mangrove Forests”, Carbon Management Iss. 3:3, pp. 312-322, 2012
3. J. B. Kauffman, A. F. Bernardino, T. O. Ferreira, L. R. Giovannoni, L. E. D. O. Gomes, D. J. Romero, L. C. Z. Jimenez, and F. Ruiz, “Carbon stocks of mangroves and salt marshes of the Amazon region, Brazil”, Biology Letters, vol. 14, no. 9, p. 20180208, 2018
4. Zegegy, Christian, Sergey Ioffe, Vincent Vanhoucke, and Alexander A. Alemi. “Inception-v4, “Inception-resnet and the impact of residual connections on learning.” AAAI conference on artificial intelligence, vol. 30, 2017
5. Xiang-Yang Wang, Ting Wang and Juan Bu, “Color image segmentation using pixel wise support vector machines”, Pattern Recognition, Vol. 44, Iss. 4, pp. 777-787, 2011

6. R. Devadas, R. J. Denham and M. Pringle, "Support Vector Machine Classification of Object-based data for Crop Mapping, using Multi-Temporal Landast Imagery", international Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences, Vol. B7, 2012
7. Maggiori, E., Y. Tarabalka, G. Charpiat, and P. Alliez, "Convolutional Neural Networks for Large-Scale Remote-Sensing Image Classification", IEEE Transactions on Geoscience and Remote Sensing, pp.645 - 657, 2017
8. Hongsheng Zhang, Luoma Wan, Guanghui Lin and Hui Lin, "A small-patched Convolutional Neural Network for mapping at species level using High Resolution remote-sensing Image", Annals of GIS, Vol. 25 , pp. 45-55, 2019
9. David Berthelot, Nicholas Carlini, Ian Goodfellow, Nicholas Papernot, Avital Oliver, Colin Raffel, "MixMatch: A Holistic Approach to Semi-Supervised Learning", ArXiv 1905.02249, 2019
10. Van Engelen and J.E.Hoos, "A survey on Semi-Supervised Learning", Mach Learn 109, pp.373-440, 2020
11. Jizong Peng, Guillermo Estrada, Marco Pedersoli, Christian Desrosiers, "Deep Co-training for Semi-supervised Image Segmentation", arXiv 1903.11233v3, 2019
12. Hao Wu, Saurabh Prasad, "Semi Supervised Deep Learning using Pseudo Labels for Hyperspectral Image Classification", IEEE Transactions on Image Processing, pp.99, 2017
13. Jian Yao, Zhonfei Zhang, "Semi Supervised Learning based Object Detection in Aerial Imagery", Computer Vision and Pattern Recognition, IEEE Computer Society Conference, 2005
14. P. Lassalle, J. Inglada, J. Michel, M. Grizonnet, and J. Malik, "A scalable tile-based framework for region-merging segmentation," IEEE Transactions on Geoscience and Remote Sensing, 2015.
15. T. Moranduzzo, M. L. Mekhalfi, and F. Melgani, "LBP-based multiclass classification method for UAV imagery," in 2015 IEEE International Geoscience and Remote Sensing Symposium (IGARSS), 2015.
16. Mohammad Peikari, Sherine Salama, Sharon Nofech-Mozes, Anne L. Martel, "A Cluster-then-label Semi-supervised Learning Approach for Pathology Image Classification",www.nature.com/scientificreports,2018
17. James Newling, François Fleuret, "Nested Mini-Batch K-Means",Conference on Neural Information Processing Systems, 2017
18. Dong-Hyun Lee. Pseudo-label : The simple and efficient semi-supervised learning method for deep neural networks. ICML 2013 Workshop : Challenges in Representation Learning (WREPL), 07 2013
- 19.Olga Zoidi, Eftychia Fotiadou, Nikos Nikolaidis, and Ioannis Pitas, "Graph-based label propagation in digital media: A review". ACM Comput. Surv. 47, ,2015
20. Leto Peel, "Graph-based semi-supervised learning for relational networks",arXiv 1612.05001 ,2016
21. Leo Breiman, "Random Forests", Statistics Department University of California Berkeley, 2001