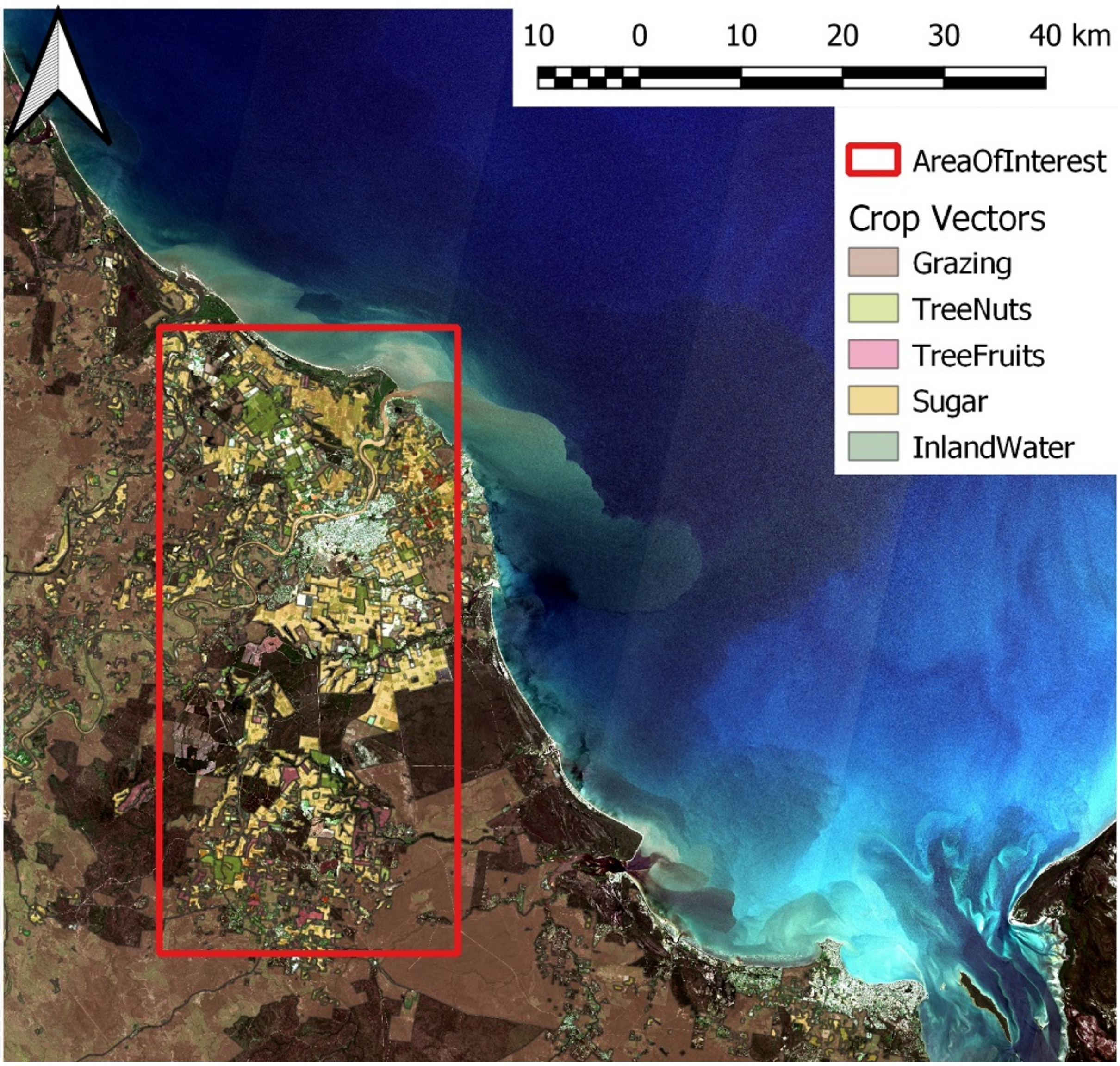
Topic 0 – Crop Classifier from Satellite Imagery

COMP9417 - Crop Classifier



Made By:

Christopher Sheaffe – z5177424

Lilian Zhou – z5131082

Sankalp Shukla – z3462149

Table of Contents

1. Introduction

There has been ongoing research into the field of crop classification through the use of machine learning. Now, with the abundance of remote sensing satellite imagery the need for classification of different pieces of land is greater than ever. Classification of crops can provide us with various insights about a particular area of land that were previously harder to obtain. This report discusses the different machine learning algorithms that were used and optimised to make this classification easier and faster. The algorithms were tested against several different images that were taken from the Sentinel-2 satellite. The images used have been taken in 2017 during different times of the year. For training and testing, satellite images of the Bundaberg region in Queensland with the dimensions of 110km x 110km were used. Two different methods that were used to form the results were ensemble learning and CNN tensor flow. The test and training accuracies of the two models was compared.

2. Research

In the last few years, several research papers have been published describing how crop classification can be done using machine learning software. One such research paper outlined the test performance of the different machine learning algorithms in crop classification. The results showed that the random forest algorithm was the best out of the bunch in terms of efficiency and accuracy. The test results obtained an accuracy of 94.6%. This was then followed by decision tree classification (DTC) with an accuracy of >90% which classified 2 crops, corn and soybean. SVM returned an accuracy of >86% in discriminating between the various crop types. [1]

Another research, in Remote Sensing of Environment journal developed a deep learning neural network algorithm that was built with one-dimensional convolutional (Conv1D) layers and tested the model against other popular crop classification algorithms such as the XGBoost, Random Forest and Support Vector. Through the use of various visualization techniques, notably NDVI, they concluded that the deep learning neural network was far superior. This was due to the inherent nature of Conv1D layers to better capture the temporal patterns and also due to their ability to handle shapes in a hierarchical manner. The Conv1D layers also automatically extract features from the training dataset which reduces the need to optimise features later on in the process. [2]

To further test and confirm the findings of the two research papers discussed above, a common test and training dataset were made, and the models were compared against each other. The models that were used in this test include the Decision Tree Classification, Bernoulli Naive Bayes, Random Forest, Ada Boost, Multi-Layer Perceptron and Ensemble Learning. Table 1 compares the precision, accuracy, recall and F1-score of these models. As it can be seen from the table that random forest is a far superior model for this particular problem.

Table 1 – Test and training accuracies of different models

Model	Accuracy	Precision	Recall	F1-Score
Decision Tree	91.48%	91.50%	91.48%	91.47%
Naïve Bayes	78.89%	62.24%	78.89%	88.20%
Random Forest	93.50%	93.38%	93.50%	93.69%
Ada Boost	86.73%	86.02%	86.73%	87.80%
Multi-Layer Perceptron	78.89%	62.24%	78.89%	88.20%
Ensemble Learning	88.14%	89.32%	88.14%	89.93%

3. Implementation

The ensemble learning, and the CNN algorithm were compared against one another. To tackle this, some of the initial steps were same however, the datasets had to be analysed separately. In order to do this, the initial problem was first subdivided into eight sections so that each of the sections can be tackled separately. The eight subdivided sections are listed below:

1. Collecting Source and Land Audit Data
2. Combining Data into common CRS
3. Crop Layer Creation
4. Raster Creation
5. CSV Creation
6. Model Creation
7. Crop Classification
8. Data Visualization

A flow chart, that highlights each of the steps and then subdivides into respective steps required to do ensemble learning and/or CNN is shown in Figure 1.

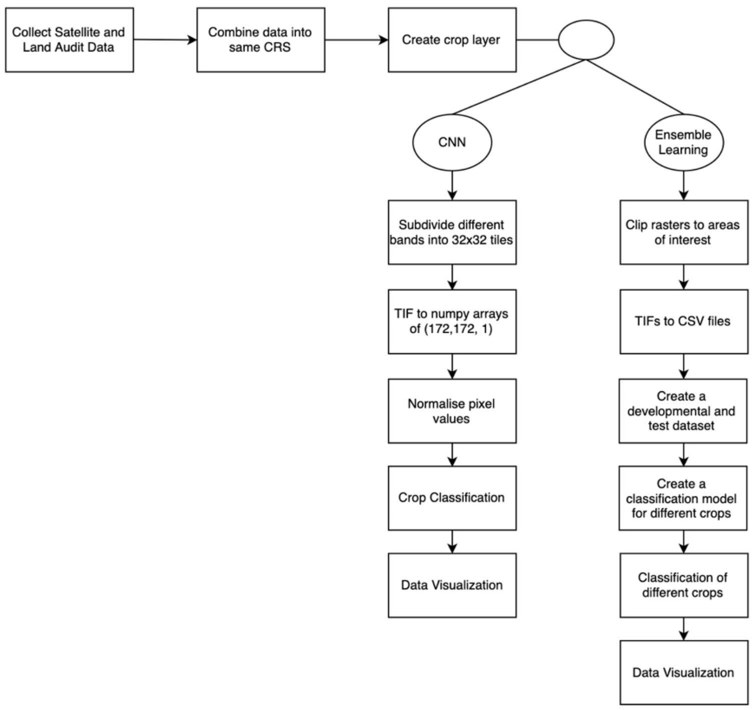


Figure 1 – Flow chart

The initial sentinel-2 imagery data was collected from the sentinel-2 website at a 10m, 20m or 60m ground resolution for the different bands and land audit data was collected from the Queensland government website. All the different bands were readjusted to a ground resolution of 20m, which has a negligible effect on the accuracy of these images. However, an error that is introduced during this step is the accuracy of the land audit data is not within the 20m ground resolution of the images. The satellite imagery contained 13 different spectral bands VNIR (Visible/Near Infrared) to SWIR (Short Wave Infrared) and an image from the visible spectrum. Bands 5, 6, and 7 were used for the CNN algorithm as these bands are the vegetation bands. All 13 bands were used to differentiate crops in the ensemble learning algorithm. Random forest used red band 4, vegetation bands 5, 6 and 7 and near infrared bands 8 and 8A. Data from different months and seasons in 2017 was collected such that the algorithm accuracy could be thoroughly tested.

For the ensemble learning section, the collected land audit and the imagery data had to be transformed into a common CRS (Coordinated Reference Systems). EPSG: 32756 was used as the common transformation system and the transformation was done using QGIS. Before this can be done however, the geometries of the vector images had to be fixed. This resulted in some of the polygons being removed which in turn caused some of the positive classes to be labelled as negative. This removal of geometries may have led to an increased number of false positives in the data. The land audit data then had to be converted into image readable vector data. This conversion helped in classifying the different crop layers on the map. This conversion should not affect the accuracy of the data or the models.

For the CNN part of the project, the satellite imagery was divided into 32x32 tiles for faster learning and testing. The tiles were then cross referenced with the crop layers to further subdivide them into areas that contained a crop and areas that did not. The pixel values then had to be normalised to binary (0s and 1s) from the initial 0 to 4095. For ensemble learning, per pixel machine learning was used and hence, the raster files of the different bands had to be converted into algorithm readable csv files. Data visualisation was done for both the algorithms and is further discussed in the Experimentation and Results section of the report. The csv files that were obtained were converted back to TIF files for better visualisation.

4. Experimentation and Results

The CNN model was trained and tested on a recommended 80% to 20% split respectively. Initially, it was trained on a relatively small dataset of 800 samples then tested on a dataset of 200 samples. To avoid overfitting different techniques such as using a relatively small architecture and multiple dropout layers were implemented. This simple architecture was known to work with the MNIST dataset was used to determine the initial accuracy. After rigorous tweaking, it was found that 5x5 filter size and a max pooling size of 2x2 provided optimal results. The performance of this dataset across the different bands was similar with a test accuracy of around 85%, 89% and 78% across bands 5, 6 and 7 respectively. This low accuracy for band 7 may have been because the model stopped learning prematurely due to finding a local minimum at a value of around 78%. The accuracies at different months and different bands are shown in Table 2 below. Band 6 performs the best with low test and training samples.

Table 2 – Test accuracies for smaller test and training datasets for CNN

Month	**Band 5**	**Band 6**	**Band 7**
March	86.83%	90.78%	82.52%
June	81.73%	90.78%	78.64%
August	87.86%	84.47%	74.27%
Average	85.47%	88.68%	78.48%

A larger dataset that spanned across the different months in 2017 was then used to train and test the model. Training set of 2400 samples and a test set of 600 sample was used again with an 80% to 20% split. As expected, the training and test data returned an accuracy of 99% and 96% respectively. This is due to increasing the number of samples at low dataset counts generally results in a greater accuracy improvement. The training set was run over 20 epochs as this was found to be the best in terms of the time taken and the final training accuracy stabilisation. The accuracy of the CNN overtime is shown in Figure 2 of this document. The relatively close accuracy values for the test and training sets suggest that there was no overfitting and hence, no further modifications were done to the model. Since the results for the three vegetation bands were similar, this model was used in the remaining datasets.

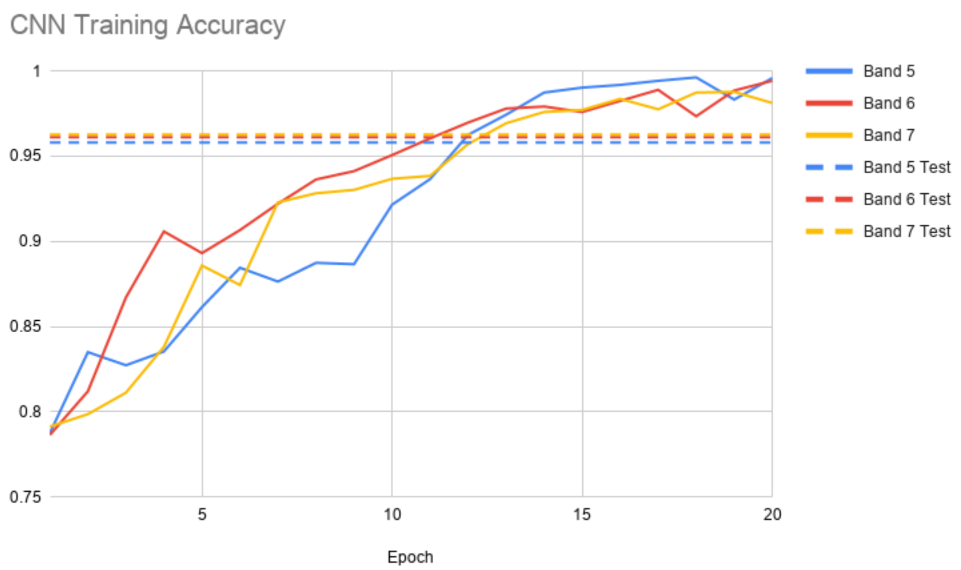


Figure 2 – CNN accuracy over 20 epochs

The test results for the three different months were all very similar and close to an accuracy of 96%. It is possible that there was little change over the three-month period we used to train the model. The model succeeds in correctly classifying the agricultural land that had similar features. As a lot of tiles were ocean and/or cloud layers, the accuracy of the model may be a bit inflated. Some of the examples that the model correctly predicted are shown in the figures below. The 3 convolution layers of band 5 of tile 192 is shown in Figure 3 to Figure 5. The original image is shown in Figure 6. This tile is a positive dataset and contains the sugarcane crop. Tile 257 is a negative dataset has an absence of the sugarcane crop. The 3 convolution layers for this tile are shown in Figure 7 to Figure 9 and the original image is shown in Figure 10. The features distinguished by the CNN can be seen in the different convolution layers. The CNN can pick up features through segmentation of layers that are not distinguished by the NDVI images. The NDVI images for the two tiles are shown in appendix section 8.3.

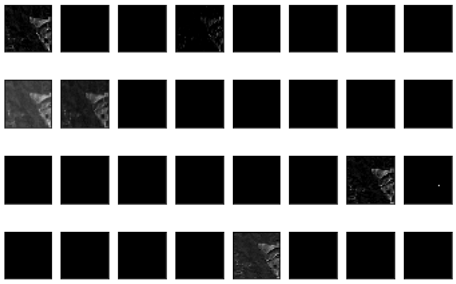


Figure 3 – Band 5 convolution layer 1 tile 192

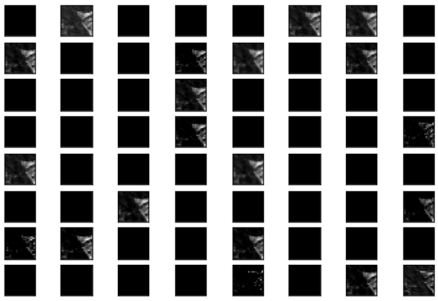


Figure 4 – Band 5 convolution layer 2 tile 192

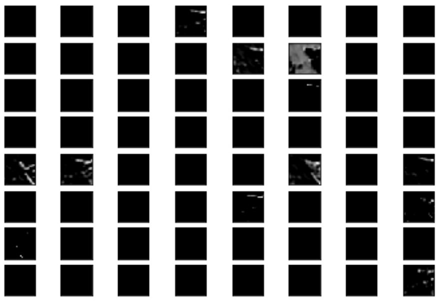


Figure 5 – Band 5 convolution layer 3 tile 192

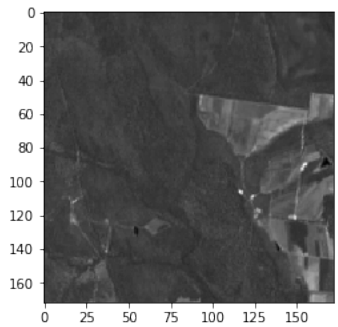


Figure 6 – Original image of tile 192

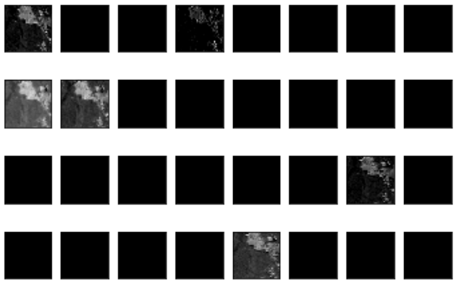


Figure 7 – Band 5 convolution layer 1 tile 257

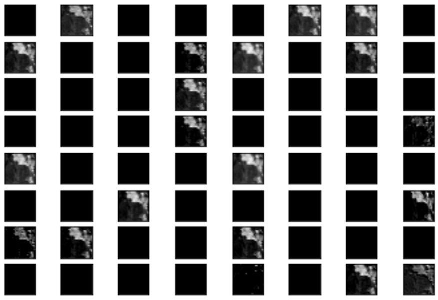


Figure 8 – Band 5 convolution layer 2 tile 257

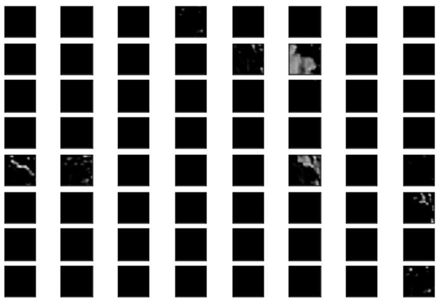


Figure 9 – Band 5 convolution layer 3 tile 257

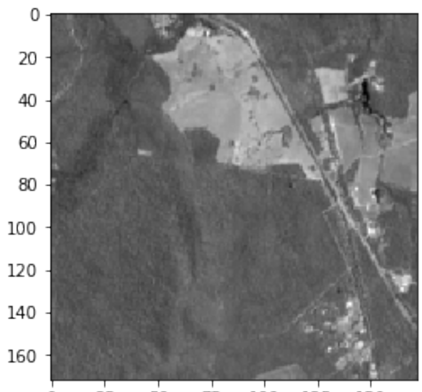


Figure 10 – Original image of tile 257

When the random forest model was created it was prone to overfitting and hence, had to be optimally modified so that the test and training accuracies did not show a large discrepancy. Some of the highly accurate results that were obtained using this model were exaggerated due to the model being able to correctly differentiate between water and land. The ocean covered almost half of the entire area which caused the model to correctly predict the results without much difficulty.

An area of extent with a large number of positive instances was used to train the random forest model such that the number of instances that were required by the training data were minimised. The area had to just be of the right size as a large area of extent resulted in the models taking far too long to train and a small area of extent resulted in underfitting which caused a decrease in accuracy as there weren't enough positive instances to train on. In order to correctly select the features for an optimal model, rigorous testing was done, and relevant features were selected. Bands outside of the Red – Near IR range were dropped and a month section was added in order to properly distinguish between different times of the year. Dimensionality reduction was not performed on this dataset.

Random forest crop classification was used on five different classes namely, sugarcane, grazing, tree fruits, tree nuts and inland water. The precision, recall, accuracy and F1 percentage scores for each of them were obtained from the developmental, test and monthly datasets have been shown in Table 3 and Table 4 The average column in Table 4 shows the average of monthly datasets in the year 2017. Month to month results are shown in appendix section 8.2 of this report.

Table 3 – Random forest accuracy scores for clipped datasets

Class	**Development Set**				**Test Set**			
	Accuracy	Precision	Recall	F1	Accuracy	Precision	Recall	F1
Sugar	99.10	99.10	99.10	99.10	79.30	77.10	79.30	77.70
Grazing	98.18	98.2	98.18	98.2	74.03	68.76	74.03	68.95
TreeFruits	99.74	99.74	99.77	99.75	98.34	96.97	98.34	97.58
TreeNuts	99.67	99.65	99.65	99.66	97.56	97.33	97.56	97.38
InlandWater	99.73	99.73	99.73	99.73	94.29	95.99	94.29	95.04

Table 4 – Random forest accuracy scores for unclipped dataset

Class	**Average**			
	Accuracy	Precision	Recall	F1
Sugar	96.26	95.81	96.26	95.97
Grazing	83.67	80.71	83.66	79.94
TreeFruits	99.68	99.48	99.68	99.55
TreeNuts	99.29	99.44	99.30	99.36
InlandWater	84.14	98.45	84.14	89.64

Grazing had the lowest average score in the unclipped datasets table. This can be due to the undefined physical characteristics of grazing. Grazing land is defined by humans and as such has no distinctly identifiable features which can easily mislead a machine learning algorithm in incorrectly identifying it. For better visualization of the different accuracies of the classes, the csv file that was obtained from running the algorithm was converted back into a TIF file and then a difference between the real and obtained data was taken. The results for tree nuts and sugarcane classifiers have been shown in the figures below.

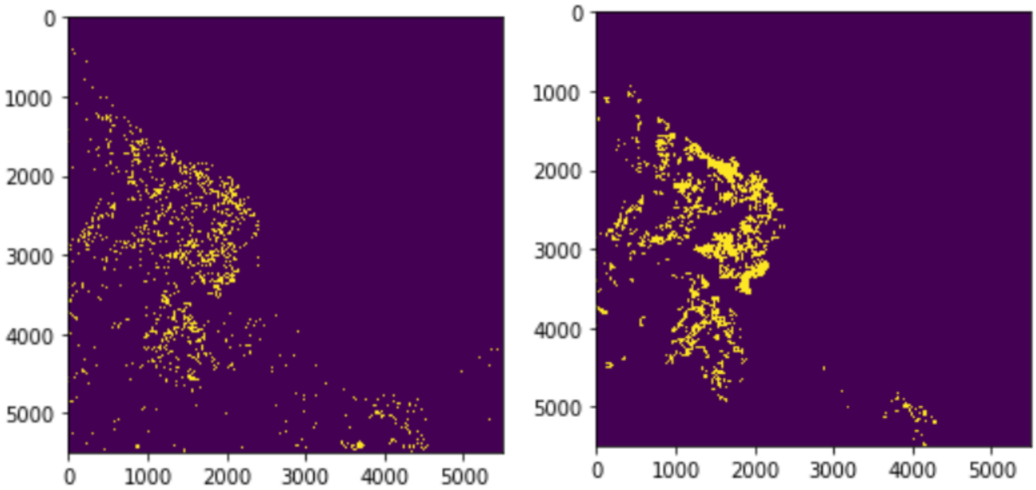


Figure 11 – Sugarcane predicted (left) and sugarcane real (right)

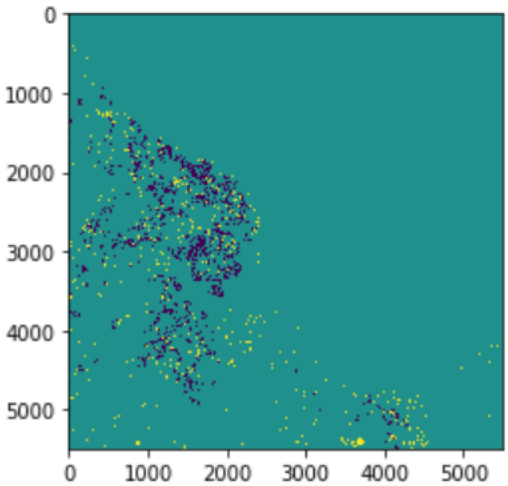


Figure 12 – Difference between sugarcane real and predicted data

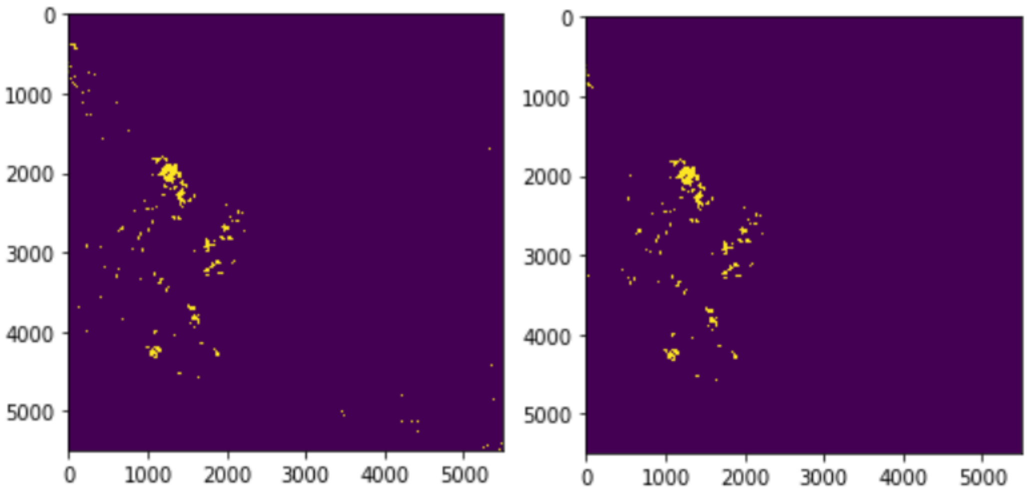


Figure 13 – Tree nuts predicted (left) and tree nuts real (right)

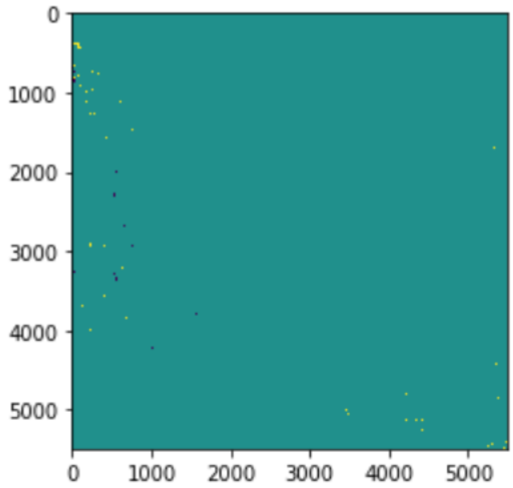


Figure 14 – Difference between tree nuts real and predicted data

5. Future Work

Some of the improvements that could be made to the ensemble learning model include the rasterization of the land audit data to 20m ground resolution. This will ensure that the land audit is the same resolution as the satellite images and will improve the accuracy of the model. Another significant improvement that could be made is finding a minimum area covered by a crop field. Doing this and then applying this minimum threshold to the model will ensure that small pixels and areas that are not large enough to be classified as a field can be ignored when displaying the results.

The CNN model can be improved in many ways, namely by removing the initial problem of local minima in small datasets. Furthermore, the model can be tested against larger datasets spanning over a larger area and during different seasons. The models built in this report did not take into account cloud cover as images with little to no cloud over positive areas were chosen. In the future, the model can be trained to disregard or incorporate cloud-like structures into its learning phase which will eliminate the need to hand-pick images containing no cloud cover.

6. Conclusion

The task of classifying the different crops on a satellite image was successfully completed and the results obtained were precise and accurate. This was finally achieved through rigorous testing and tuning of various features of the two algorithms. In conclusion, it can be said that the CNN and random forest algorithm both are best suited for crop classification. However, it is still not a certainty that these algorithms can be used on a totally different area of land.

7. Bibliography

[1] I. Ali, F. Greifeneder, J. Stamenkovic, M. Neumann and C. Notarnicola, “Review of Machine Learning Approaches for Biomass and Soil Moisture Retrievals from Remote Sensing Data,” *Remote Sensing,* vol. 7, 2015.

[2] L. Zhong, L. Hu and H. Zhou, “Deep learning based multi-temporal crop classification,” *Remote Sensing of Environment,* vol. 221, no. 0034-4257, pp. 430 - 443, 2019.

8. Appendix

8.1 Predicted and Real Data of the Rest of the Crop Classifiers

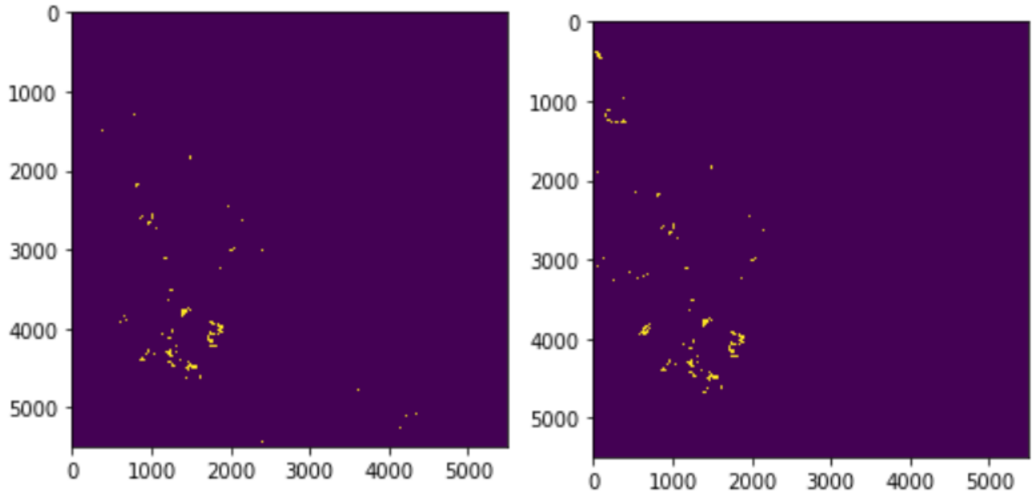


Figure 15 – Tree fruits predicted (left) and tree fruits real (right)

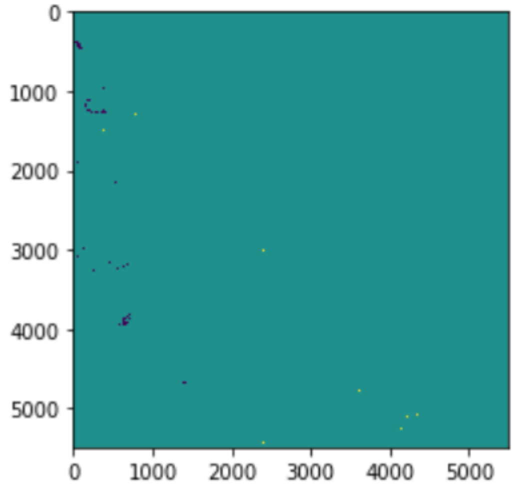


Figure 16 – Difference between tree fruit real and predicted data

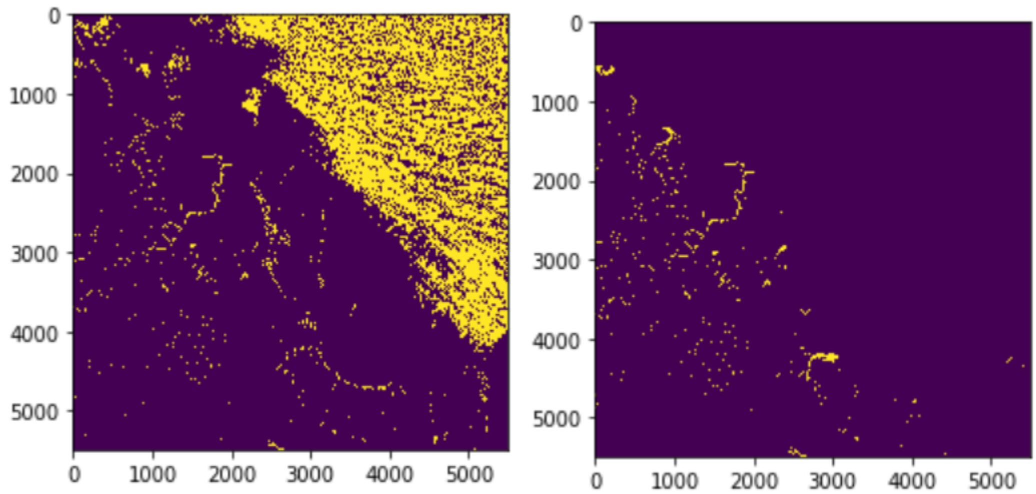


Figure 17 – Inland water predicted (left) and inland water real (right)

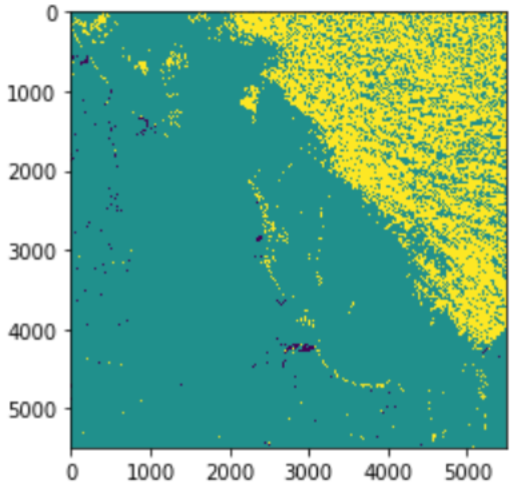


Figure 18 – Difference between inland water real and predicted data

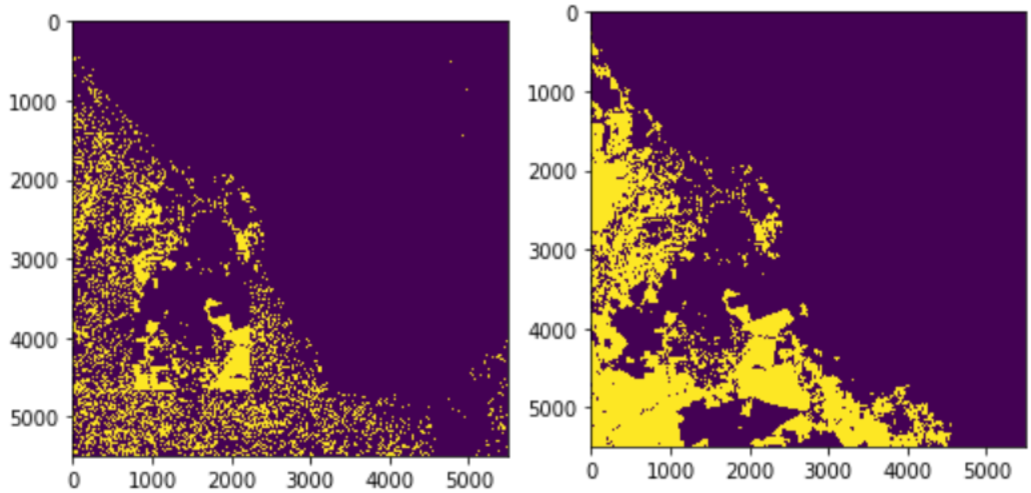


Figure 19 – Grazing predicted (left) and grazing real (right)

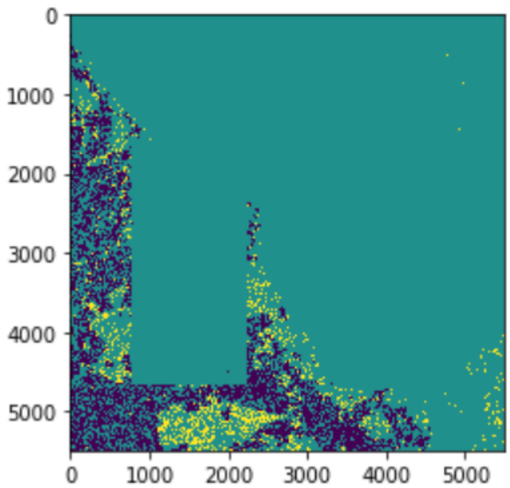


Figure 20 – Difference between grazing real and predicted data

8.2 Monthly Tables of Scores Obtained from Random Forest

Table 5 – Random forest accuracy scores on 6/3

Class/Date	6/3/17			
	Accuracy	Precision	Recall	F1
Sugar	95.95	95.28	95.95	95.54
Grazing	83.27	80.81	83.23	81.19
TreeFruits	99.63	99.36	99.63	99.48
TreeNuts	98.99	99.21	98.99	99.09
InlandWater	85.25	98.11	85.25	91.01

Table 6 – Random forest accuracy scores 24/6

Class/Date	24/6/17			
	Accuracy	Precision	Recall	F1
Sugar	95.12	94.10	95.12	94.55
Grazing	82.37	78.35	82.37	76.82
TreeFruits	99.64	99.36	99.64	99.48
TreeNuts	98.51	99.21	98.52	98.82
InlandWater	99.15	99.02	99.15	99.06

Table 7 – Random forest accuracy scores on 8/8

Class/Date	8/8/17			
	Accuracy	Precision	Recall	F1
Sugar	96.23	95.46	96.23	95.70
Grazing	82.37	78.35	82.37	76.82
TreeFruits	99.65	99.41	99.65	99.49
TreeNuts	99.31	99.35	99.31	99.33
InlandWater	90.29	98.33	90.29	93.96

Table 8 – Random forest accuracy scores on 7/9

Class/Date	7/9/17			
	Accuracy	Precision	Recall	F1
Sugar	98.73	98.83	98.73	98.77
Grazing	83.16	80.3	83.16	78.82
TreeFruits	99.65	99.5	99.65	99.5
TreeNuts	99.57	99.52	99.57	99.52
InlandWater	98.75	98.73	98.75	98.74

Table 9 – Random forest accuracy scores on 17/19

Class/Date	17/9/17			
	Accuracy	Precision	Recall	F1
Sugar	94.76	94.81	94.76	94.78
Grazing	88.35	87.56	88.35	87.33
TreeFruits	99.89	99.88	99.89	99.88
TreeNuts	99.89	99.9	99.9	99.9
InlandWater	79.43	98.54	79.43	87.58

Table 10 – Random forest accuracy scores on 27/10

Class/Date	27/10/17			
	Accuracy	Precision	Recall	F1
Sugar	96.76	96.35	96.76	96.48
Grazing	82.5	78.87	82.5	78.65
TreeFruits	99.64	99.34	99.64	99.48
TreeNuts	99.48	99.46	99.48	99.47
InlandWater	51.97	97.95	51.97	67.47

8.3 NDVI Images for Tiles 192 and 257

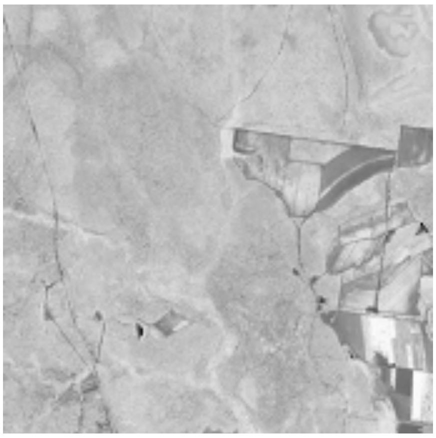


Figure 21 – NDVI image for tile 192

Figure 22 – NDVI image for tile 257