Drone Crop Image Analysis

B. Tech Project

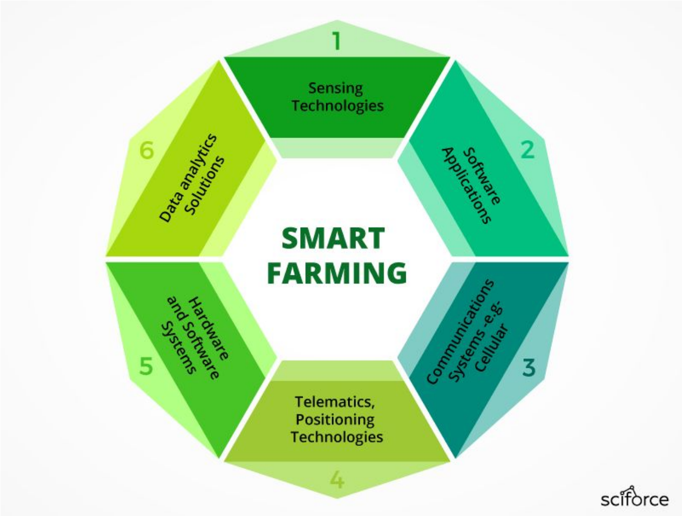
Ameya Vadnere, Ojas Raundale
170010002, 170010004

Advisor: Prof. Gayathri A., IIT-Dh

Introduction

Smart Agriculture

With the boom of IT and AI in the last decade, there has been an increase in the use modern technology to assist farming.



Smart Farming

Some Technologies that are currently available to farmers:

- Sensing technologies, including soil scanning, water, light, humidity, temperature management;
- Software applications — specialized software solutions that target specific farm types;
- Communication technologies, such as cellular communication;
- Multispectral/Hyperspectral Imaging

Multispectral Imaging

- Multispectral Imaging can help us in determining plant health over large areas, which can help in cutting down losses. This can otherwise consume a lot of time.

- Classification of different types and species of crops can help in determining/monitoring agricultural produce over a region.

- This type of classification will be our focus in this project.

Multispectral Imaging

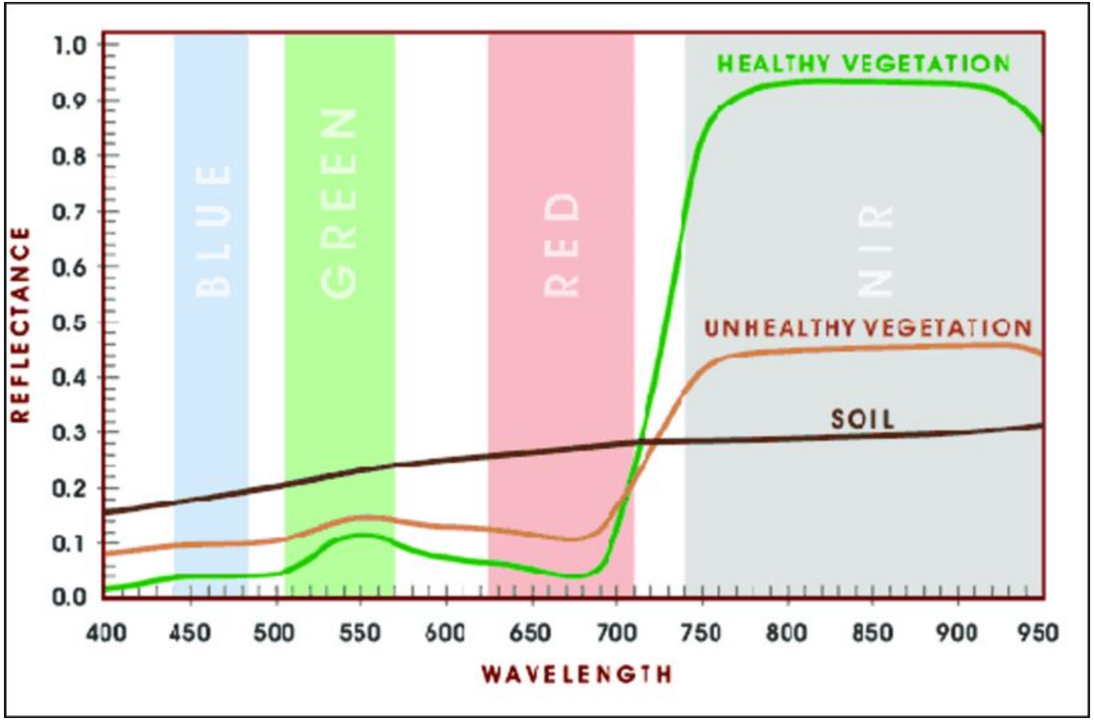
- Traditionally, RGB images have been used for classification for a long time.

- However, it is now known that sensing objects' data in some bands/wavelengths that are indiscernible to the human eye might contain distinguishing characteristics which might help us to classify objects better.

- This applies well to vegetation in general.

- Observe that in the Near InfraRed region, i.e wavelength range of 750-1300nm, the reflectance of healthy crops is very high. This is a distinguishing characteristic.

- Different crops have different spectral signatures. Adding additional relevant bands like NIR and Rededge can provide more information.



https://www.researchgate.net/figure/Spectral-reflectance-of-healthy-plant-unhealthy-plant-and-soil-in-visible-and-NIR_fig1_299388638

Multispectral Imaging

- Multispectral imaging can also help us in calculating vegetation index (VI), which is a measure of photosynthetic activity of crops in a region.

- One such vegetation index is the Normalized Difference Vegetation Index (NDVI), which is widely used in determining plant health.

- NDVI is calculated as: NDVI = (NIR - RED)/(NIR+RED), where NIR and RED are the spectral reflectance values of the NIR band and RED band respectively. Higher the NDVI value, healthier is the crop.

https://www.pix4d.com/blog/pix4dmapper-optimizing-the-ROI-of-fungicides-with-NDVI

NDVI map. The greener regions are healthier regions. If the crop is stressed, it would show low NDVI value, prompting for timely action.

Dataset and Preprocessing

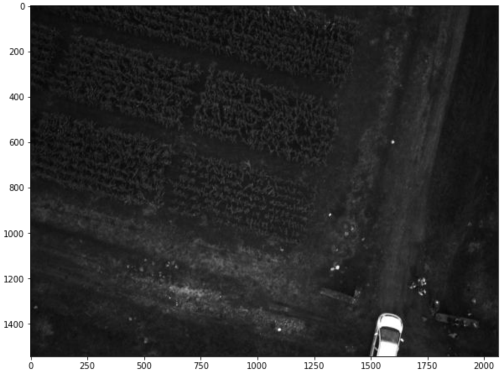
About the Data

Our data was collected over several flight missions of a rotary-wing UAV drone over an agricultural region near Dharwad in October 2019. The drone was mounted with a Micasense™ Altum Multispectral Camera.

The drone flew at an altitude of about 40m and over the flight, the camera captured high resolution (1544 x 2064) images across five bands: Red, Green, Blue, Near-InfraRed (NIR), Rededge (RE).

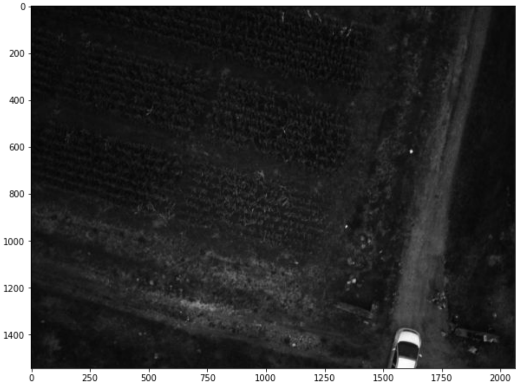
Drone

Micasense Altum



Blue

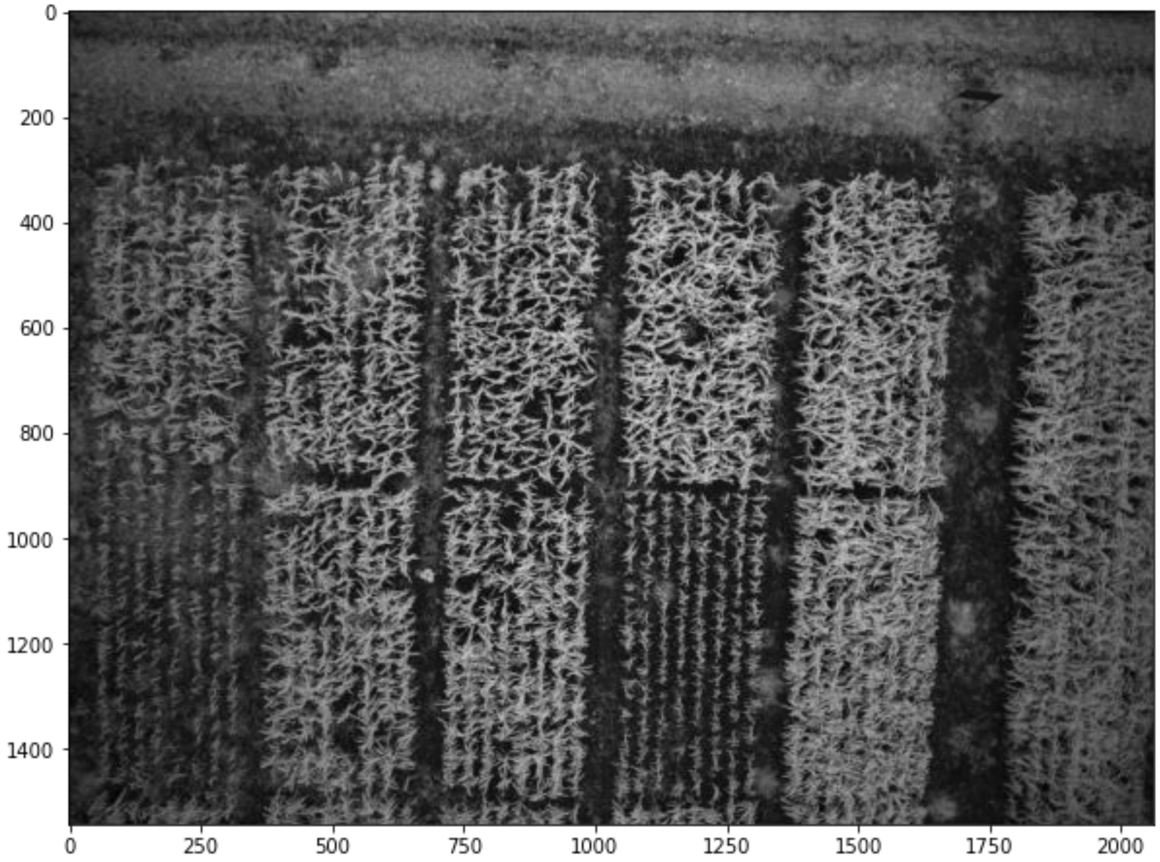
Green



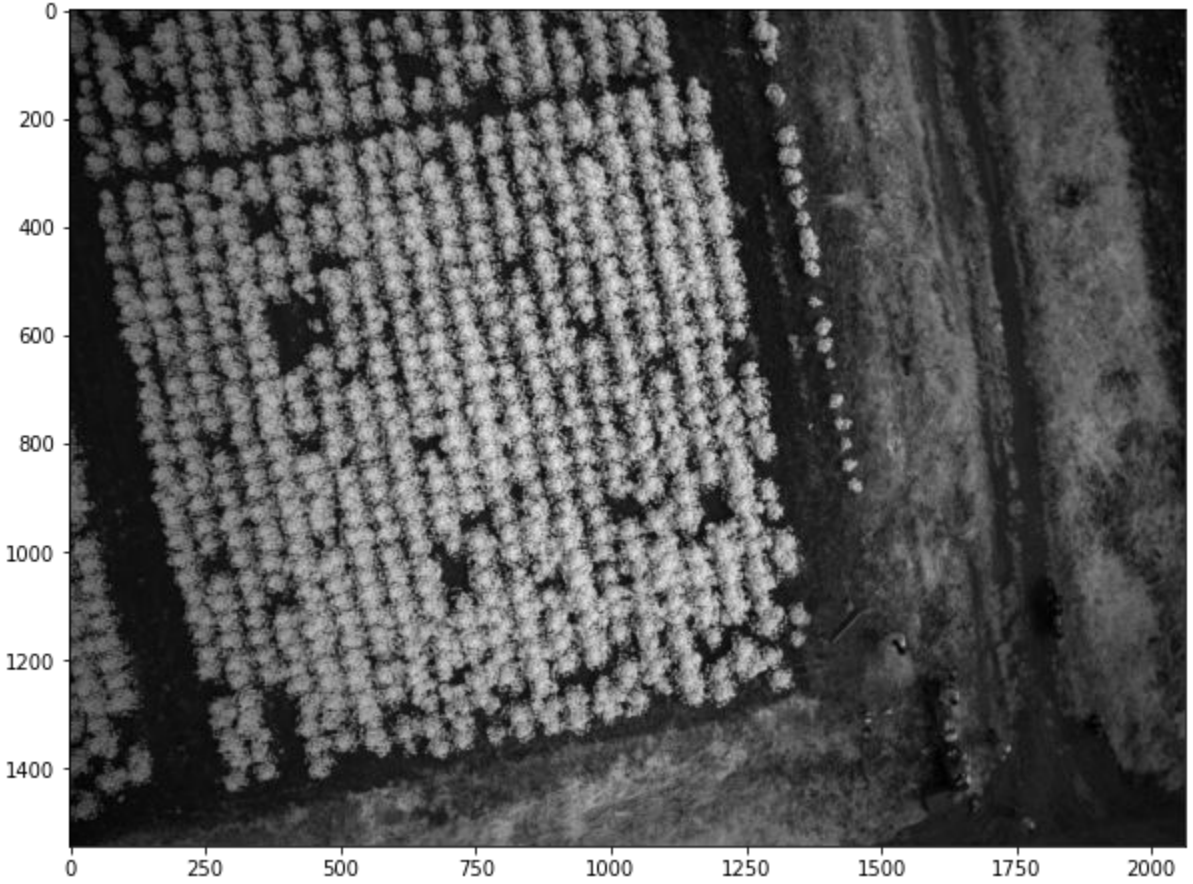
Red

Near Infrared

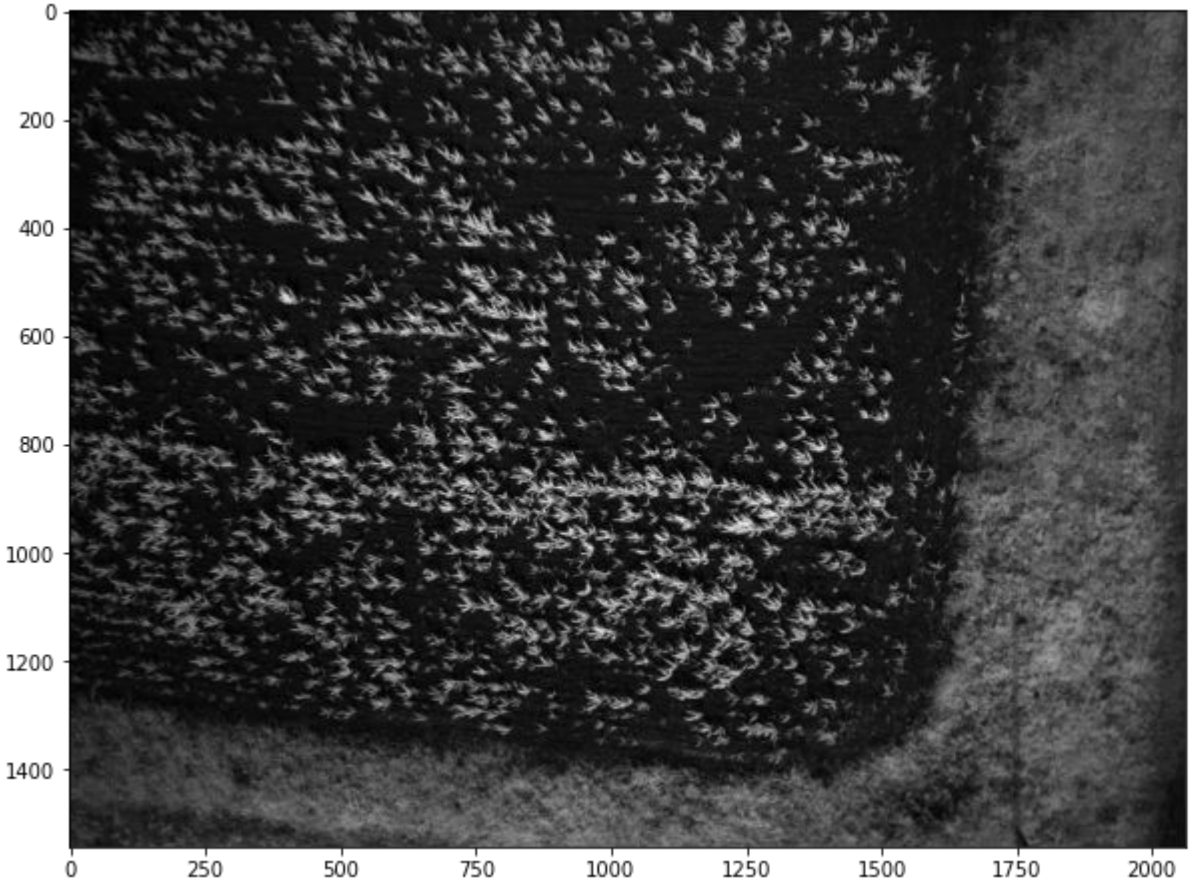
Rededge


0
200
400
600
800
1000
1200
1400
0
250
500
750
1000
1250
1500
1750
2000

Maize crop


0
200
400
600
800
1000
1200
1400
0
250
500
750
1000
1250
1500
1750
2000

Cotton crop


0
200
400
600
800
1000
1200
1400
0
250
500
750
1000
1250
1500
1750
2000

Wheat

Softwares and Applications Used

Agisoft Metashape

Agisoft Metashape (previously known as Agisoft PhotoScan) is a tool for a photogrammetry purposes, i.e creating 3D models, orthomosaics, etc.

It is widely used by archaeologists and many UAV companies. It is also used extensively in Film and Video Game industries.

We mainly use Agisoft Metashape for visualization purposes.

Agisoft Metashape

The image Dataset contains images of the field taken from different positions as the drone flies.

An **orthomosaic** (interactive drone map) is a geometrically correct aerial image that is composed of many individual still images that are stitched together. Orthomosaics provide a similar view to what you'd see in the satellite view in Google Maps.

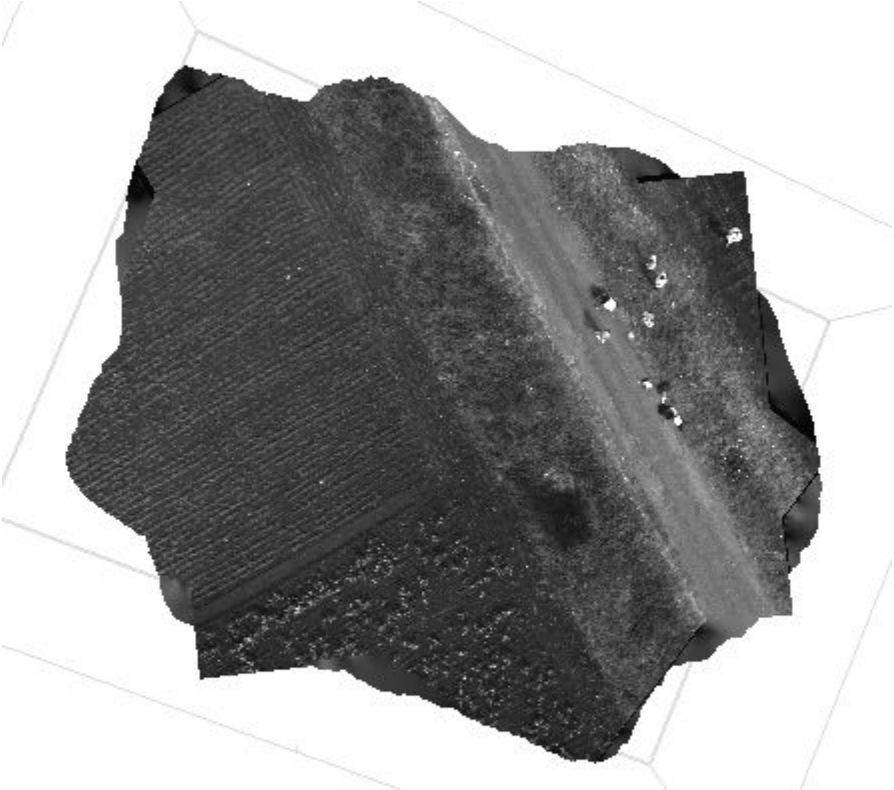
Metashape provides the feature to create a combined/orthomosaic image of all the images present in a single flight. This can give us a birds-eye view to the field covered by all the images of the same.

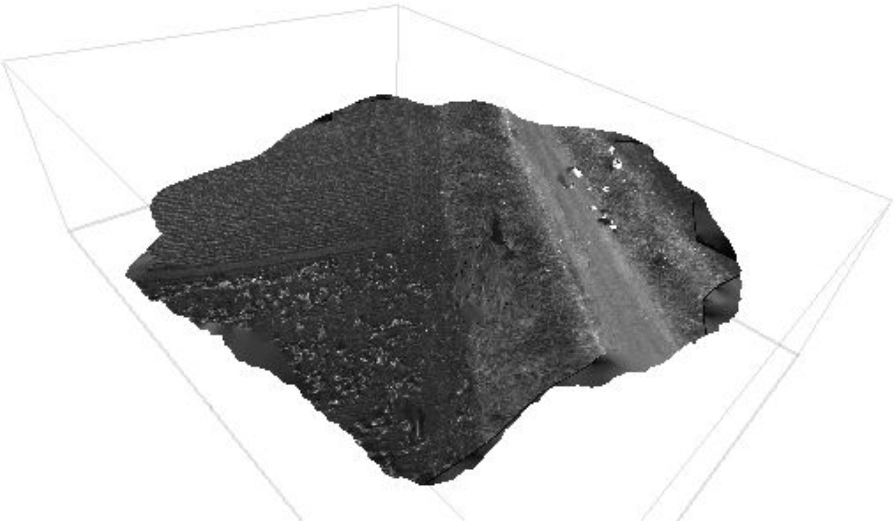
Visualizations through Metashape

Combining 4 Red band images to form an orthomosaic image

+

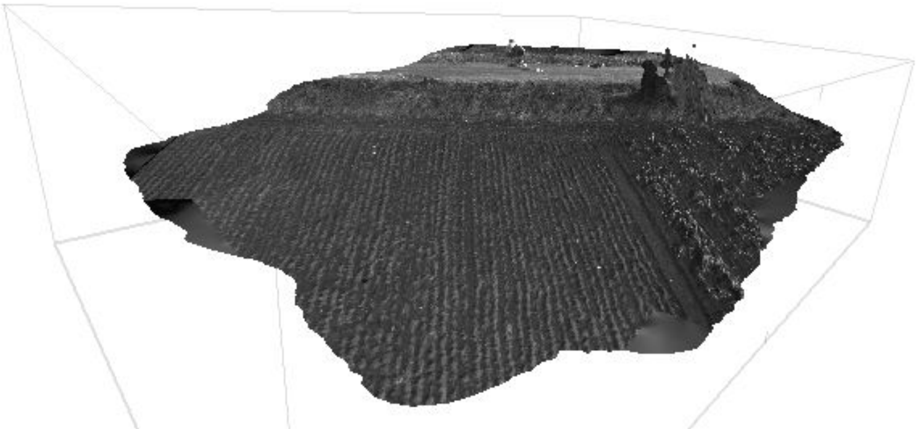
=





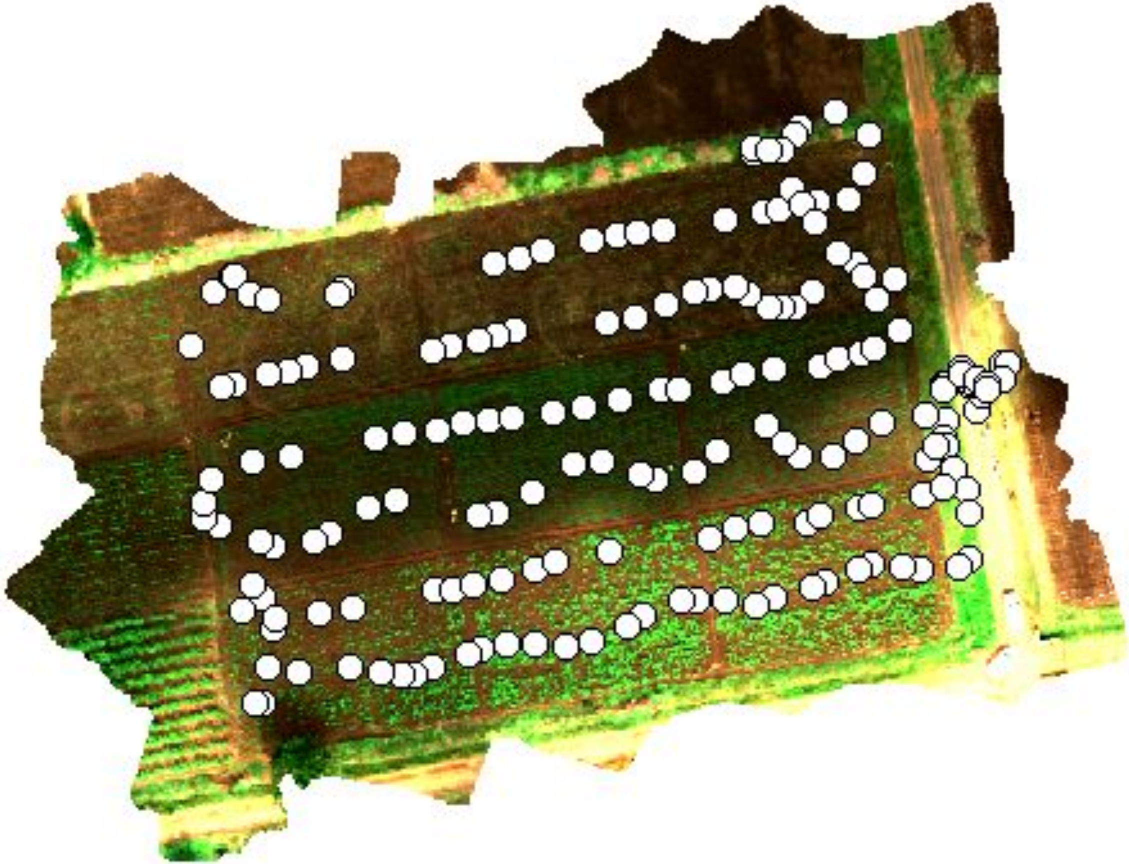
Agisoft creates a 3D Mesh of the input which can be later viewed in different angles

These 2 images are the of the same 4 images in previous slide but from different viewing angles!



The same image when combined with all the bands combined produces the following Orthomosaic image.

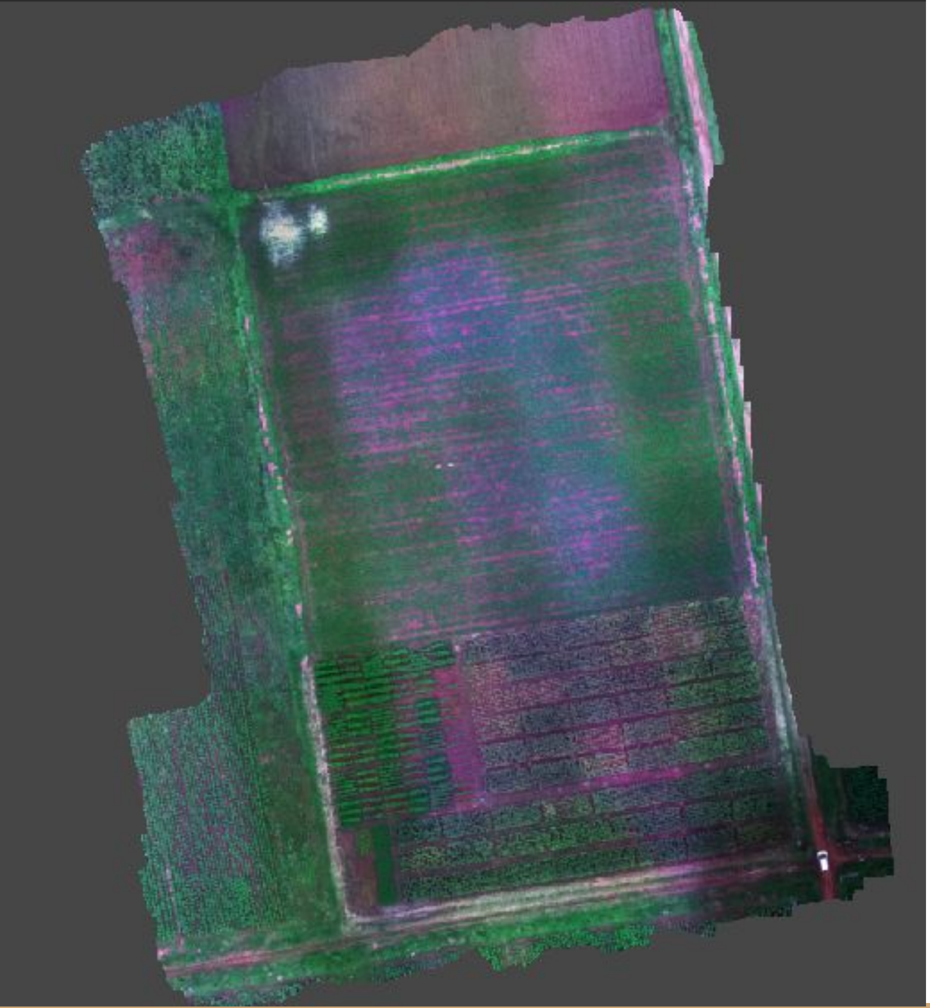
Also notice that the points at which the original image shots were taken are also shown.



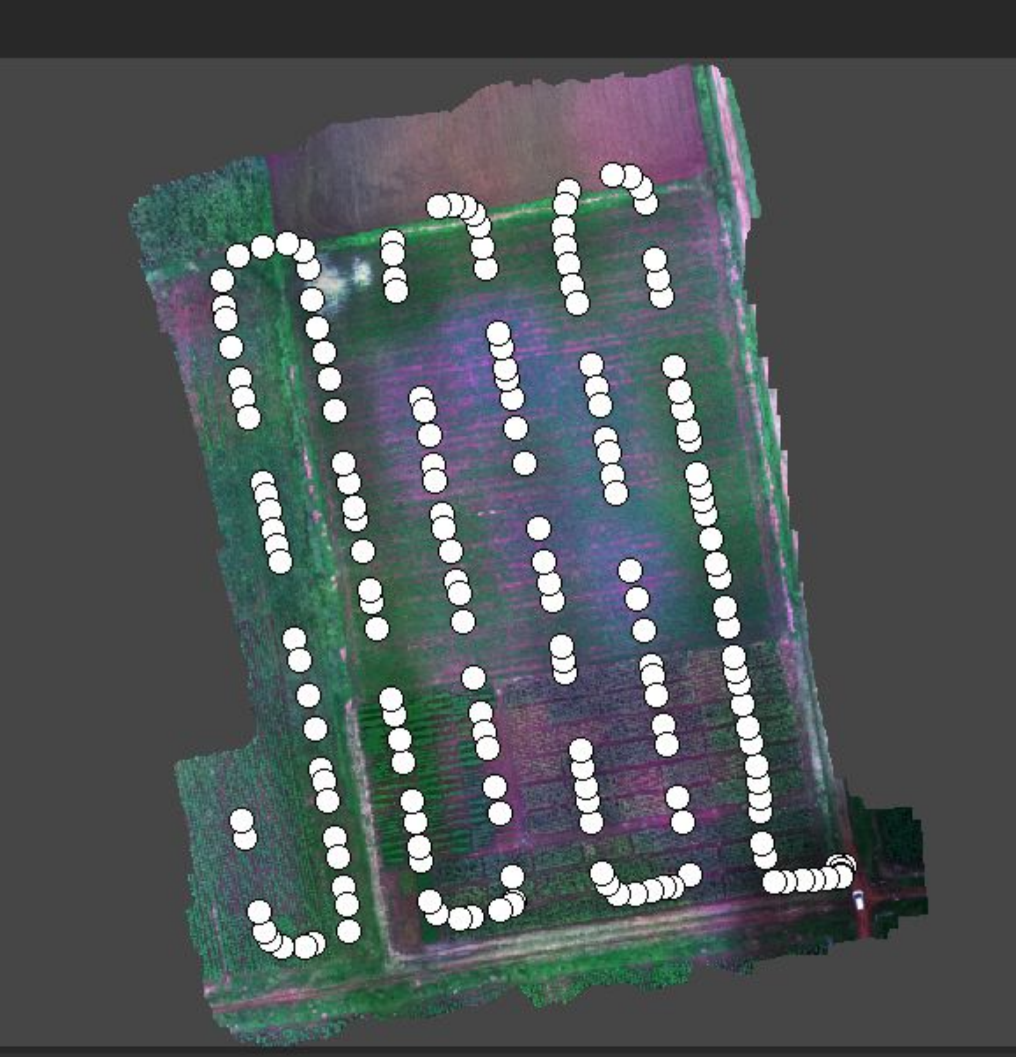
Orthomosaic of 184 images and the drone position from which they were taken

NDVI map of the orthomosaic. Darker regions show more vegetative cover.

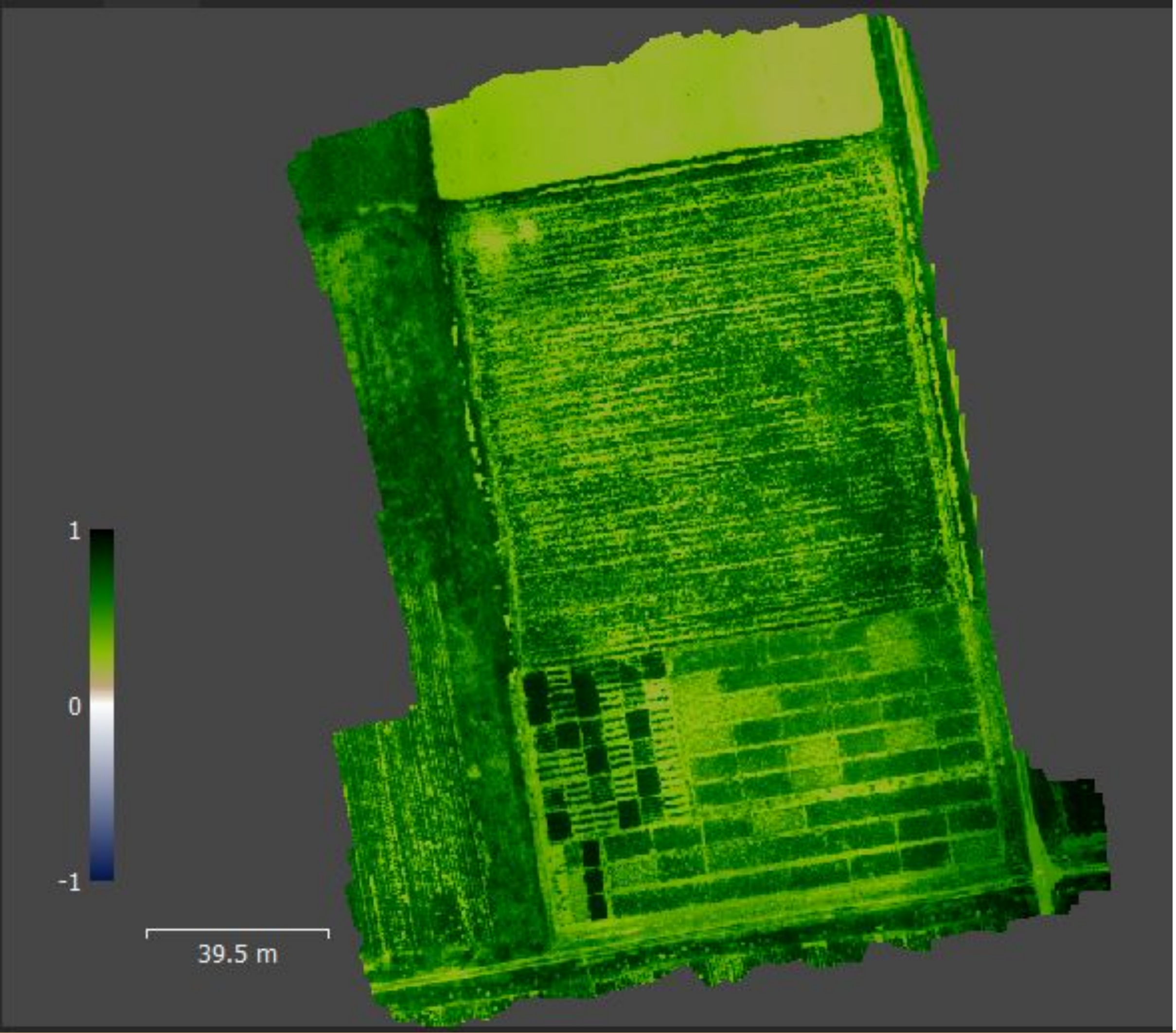
NDVI = (NIR - RED)/(NIR+RED)



This is an orthomosaic of 190 images (including all the 5 bands)



Same orthomosaic with the 190 camera positions (capture points).



NDVI indexed orthomosaic

Caveats while using Agisoft

- Agisoft, while being useful, is a commercial and premium software, with a high cost.
- The demo version does not provide sufficient functionality.
- The processing and orthomosaic creation is a computationally demanding procedure.
- Additional training is required to properly use to the software.

Micasense Image Processing library

- As discussed before, Agisoft demo has its limitations. It was not helpful for further preprocessing. We only used it for visualizations.

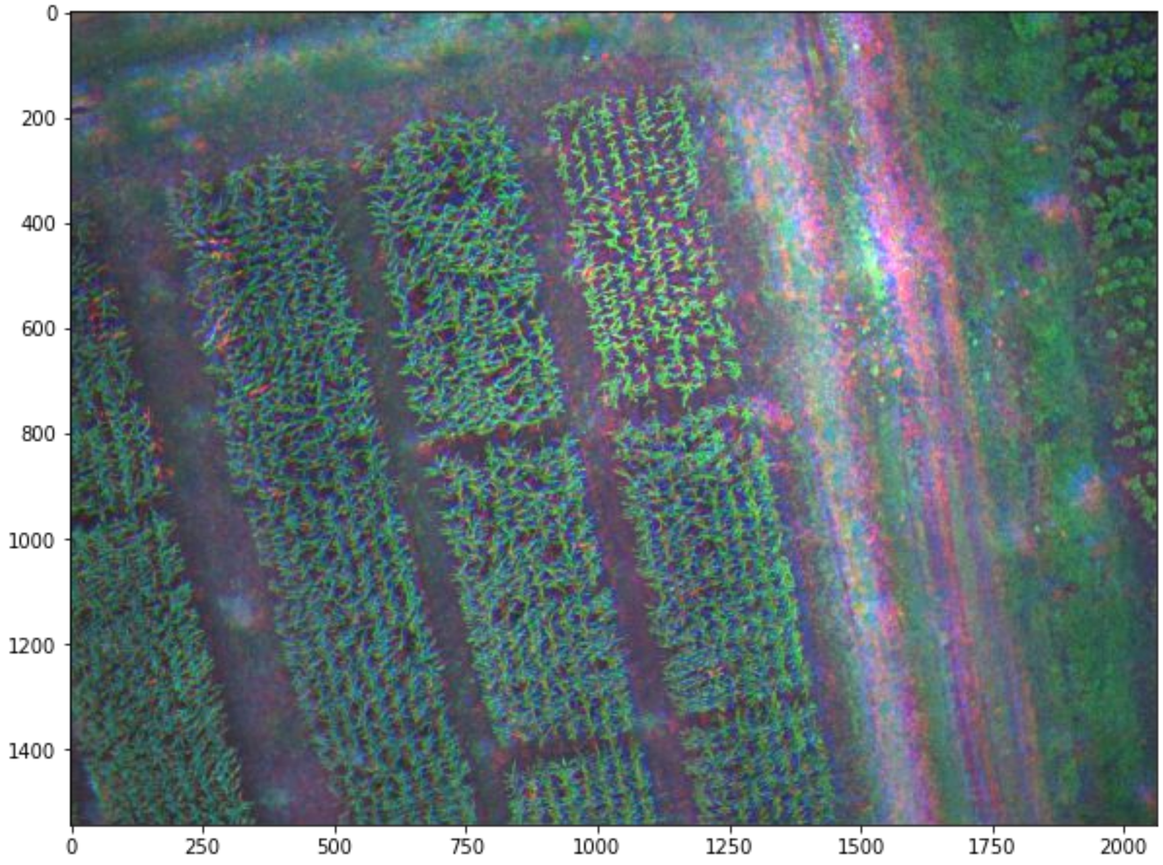
- Hence, for further preprocessing, we had to opt for an open-source library Micasense, which was developed by the manufacturers of our camera. It is built on Python.

Preprocessing

Alignment

- We have 5 different bands. We cannot stack all these bands directly and process them.
- Notice that there are 5 different lenses on the camera, so there would be some offset between the images.
- Hence, the bands of the images need to be aligned first.

If we stack RGB bands directly, the resulting image would be blurry due to the offset in bands.



Alignment

- Micasense library provides the facility to align the image with a good degree of accuracy, after which the images can be used for further processing.

- Using the metadata of the images and camera properties, Micasense finds a good transformation to align the images.

Image after alignment

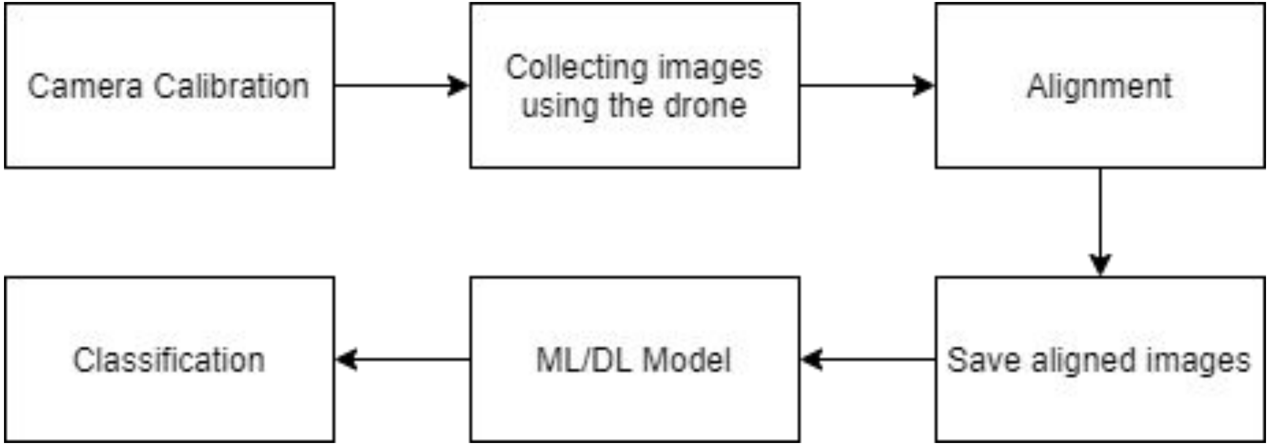
Note that this image is displayed with only the RGB colors. But the actual image contains 5 bands namely RGB, NIR and Rededge bands.

Alignment

- Thus, 5 images combine to form 1 single stack of 5 bands. These stacks of images are now ready for further processing and data analysis.

- This image alignment step was carried out to create stacks for all the available data.

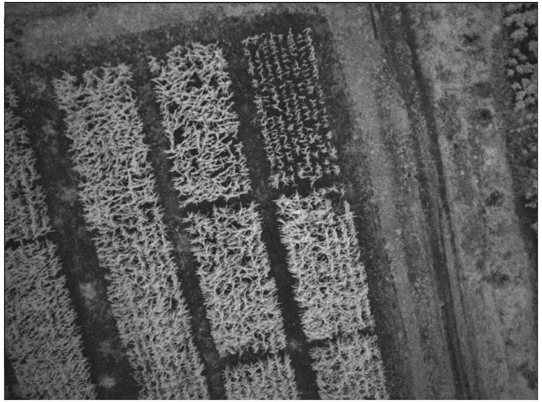
Workflow


Camera Calibration
Collecting images using the drone
Alignment
Save aligned images
ML/DL Model
Classification

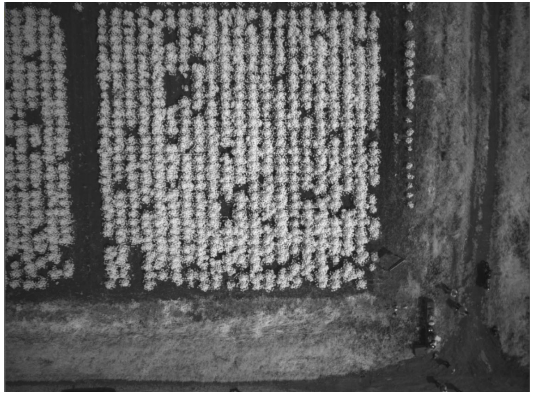
Multispectral Image Classification

Majority Classification - A simple model to test whether the data will yield results

- A model to classify an entire 1544x2064 image.
- Only using the NIR band.
- Images labelled based on the majority of crop present in the image.
- Training images used: 218 train images, 73 test images
- Class distribution : Training:85 Maize, 39 Cotton, 94 Wheat
  Testing: 29 Maize, 8 Cotton, 36 Wheat
- A basic CNN with 5 hidden layers and max-pooling yielded a train accuracy of 98.9% and a test accuracy of 95.8%



Maize



Cotton

Wheat

Some training examples for the majority classifier

Majority Classification - A simple model to test whether the data will yield results

- Note that the input images are the entire drone images. So there will be a lot of overlap of crop area within the images.
- So even if the model overfits the training set, it will still have a good performance on test data because the test data was slightly biased.
- So test dataset may not be a correct indicator generalized accuracy.
- Still, it implied that the neural network can learn how to classify crops using the NIR band.
- This was a sign for us to proceed with aligning all 5 band of the images together.

- However, assigning a label to the entire region, based on the majority crop in the image is technically incorrect and not practically useful.

- This is also due to the reason that the drone shots might consist of multiple crops in a single image.

Windowing of the aligned images

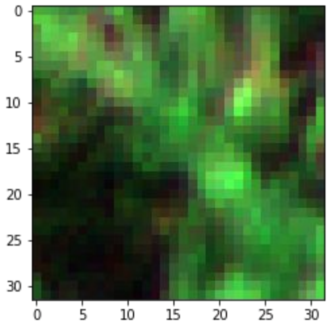
Choosing the window size

After using Agisoft Metashape to figure out how much area an average drone shot covers:

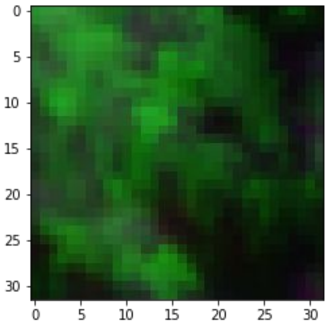
Width of area covered : 27-29 m
Length of area covered: 36-39 m
Dimension of image: 1544 x 2064 pixels

So we choose a window of 32 x 32 resolution to classify; it covers about 0.6m x 0.6m on ground, which is about the same area as that of an average crop.

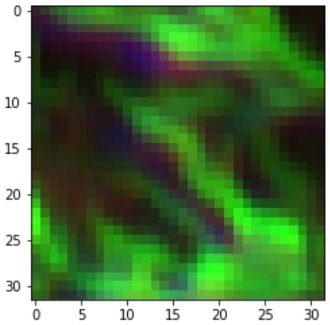
Examples of crops when the entire image is windowed by 32x32 sized window



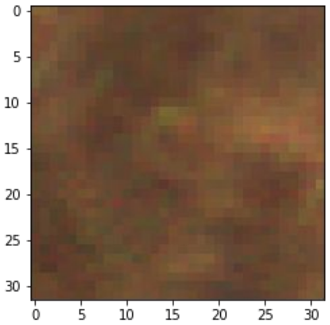
Maize



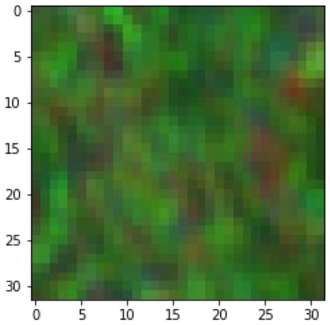
Cotton



Wheat



Soil



Grass

Windowed dataset preparation and distribution

Image Selection and Classes used

Aligned images were selected to be windowed in such a way that there is no overlap of crop area cover b/w any of them.

The following classes were chosen to be labelled and classified:

1. Cotton
2. Grass
3. Maize
4. Soil
5. Wheat

Cotton, **Grass** and **Maize** contains crops of the same.

Soil contains images of bare land cover and also the the soil present in b/w cropping areas

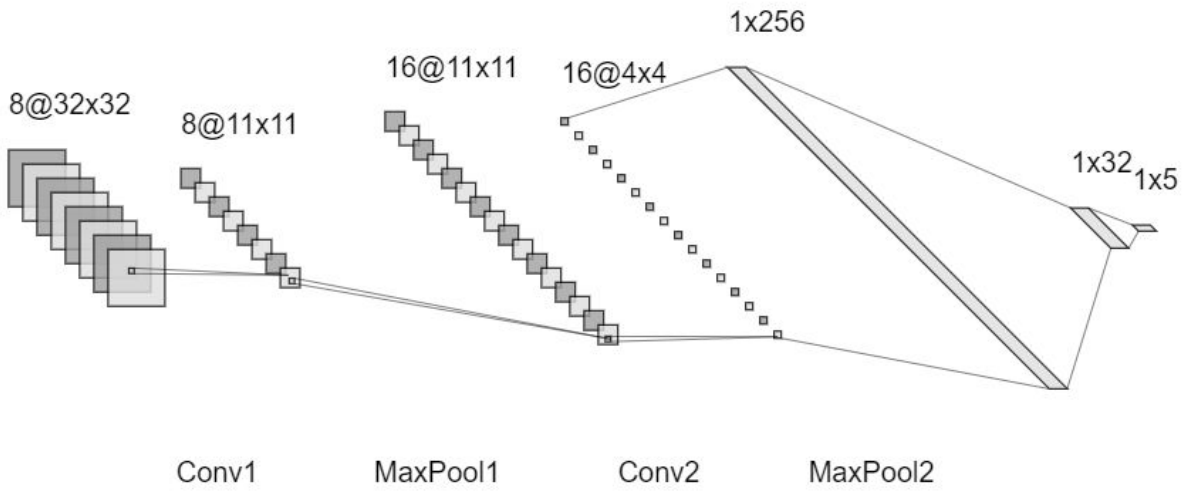
Grass contains images of unclassifiable plants and stray grass/weed

Class Distribution

Class	Aligned Images used	Total 32x32 segments	Segments used for Training	Segments used for testing
Cotton	4	2798	2395	403
Grass	8	3000	2500	500
Maize	4	3552	3123	429
Soil	7	3000	2500	500
Wheat	5	2896	2664	432
Total	**28**	**15,237**	**13,182**	**2,264**

Training the Classifiers

Model Architecture (CNN)



We experimented by tweaking some parameters and found that the following architecture works reasonably well and fast.

Models trained

Following are the models which we majorly experimented on:

1. CNN using only the NIR band.
2. CNN using RGB bands.
3. CNN using all the 5 bands.
4. CNN 5-band but without training wheat
5. Random Forest using all 5 bands
6. Random Forest 5-band but without training wheat

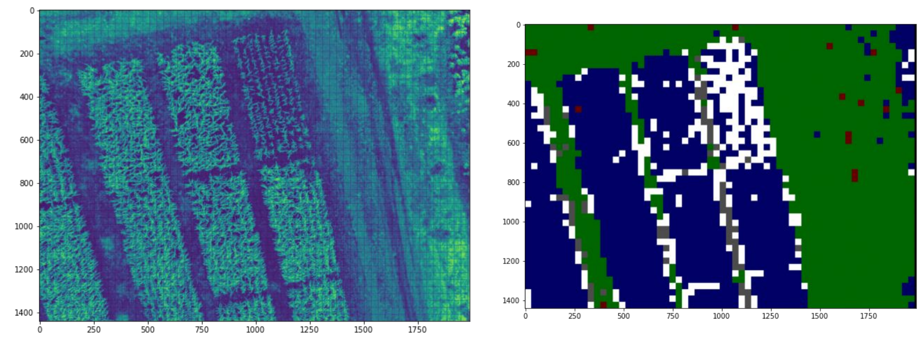
All these models were experimented on different optimizers like SGD, Adam, RMSProp, different activation functions and also different learning rates.

1.) CNN using only the NIR band

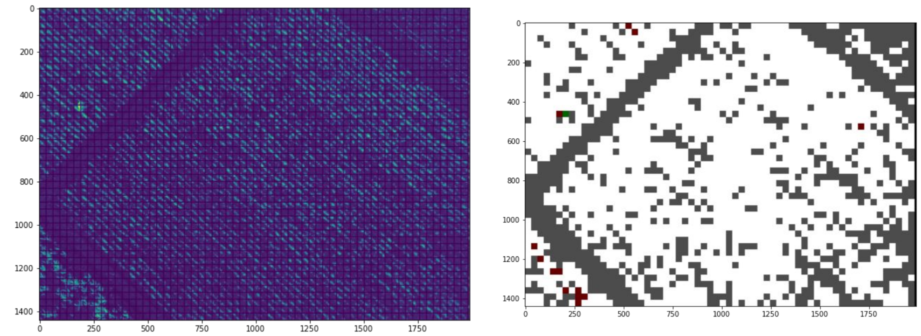
This model was trained only on the Near Infra Red (NIR) band of the train images. Following are the best results achieved:

Train	: 95.3%
Test	: 90.3%

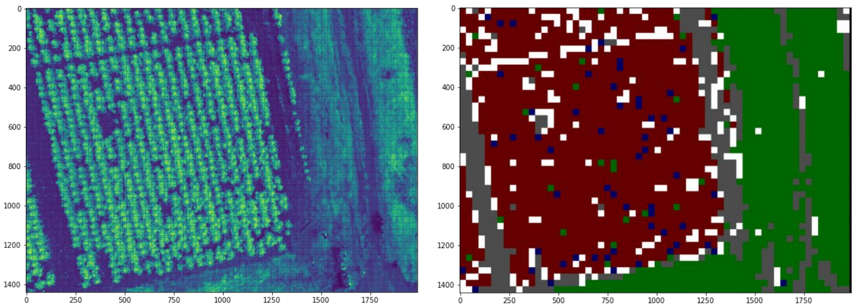
Model 1 (NIR CNN)



NIR image containing Maize



NIR image containing Wheat



NIR image containing Cotton

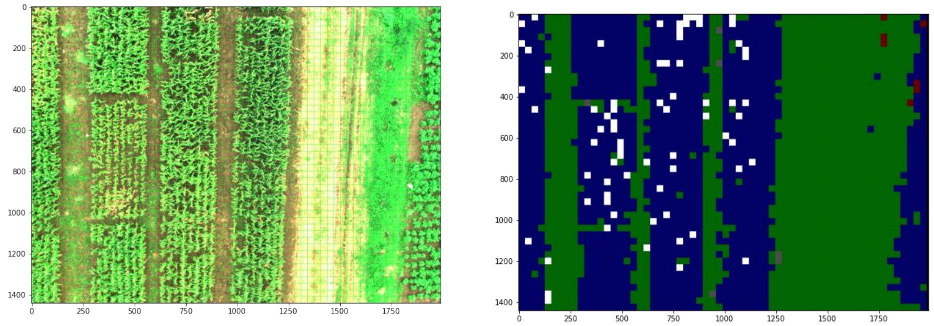
2.) CNN using RGB images

This model was trained on the RGB bands of the train images. Following are the best results achieved:

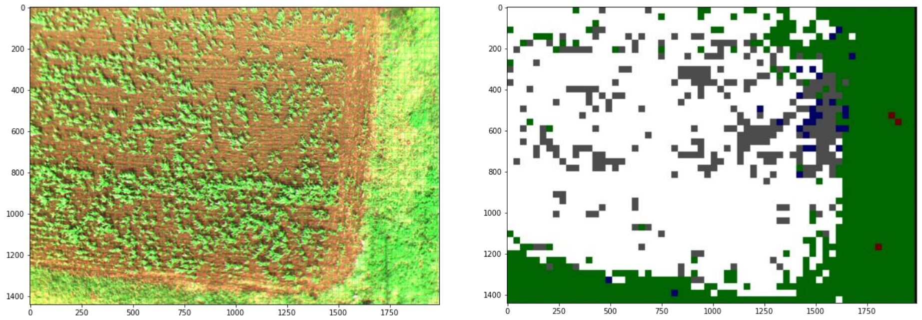
Train : 96.2%

Test : 92.8%

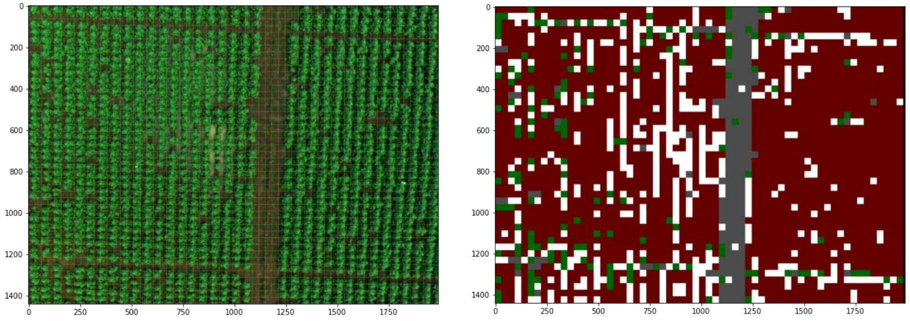
Model 2 (RGB CNN)



RGB image containing Maize



RGB image containing Wheat



RGB image containing Cotton

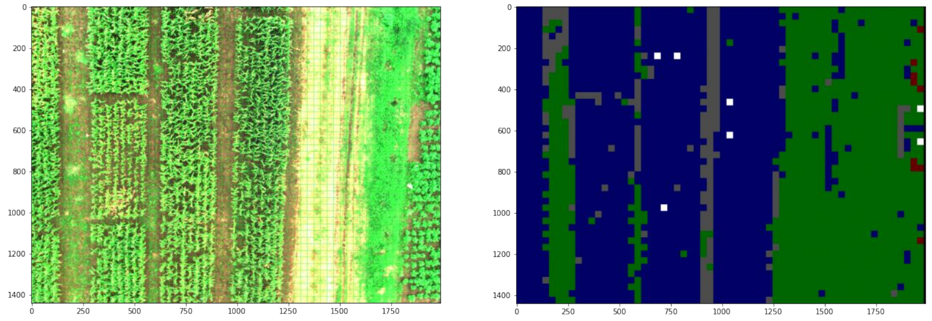
3.) CNN using all the 5 bands

This model was trained on all the 5-bands of the train images. Following are the best results achieved:

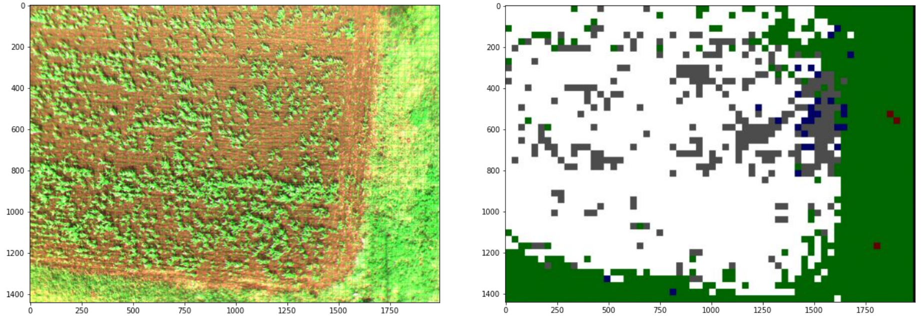
Train : 98.3%

Test : 96.68%

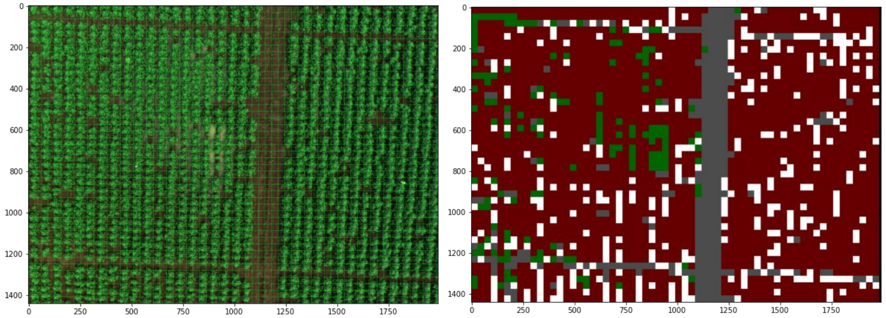
Model 3
(5-Band CNN)



5-Band image containing Maize

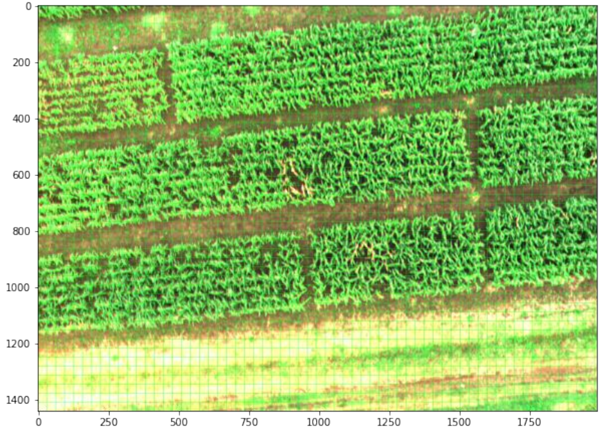


5-band image containing Wheat

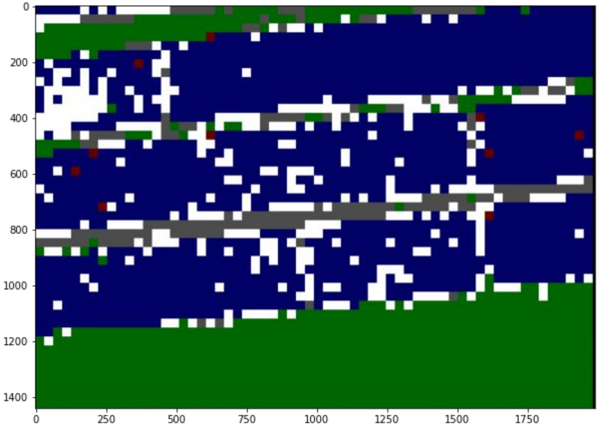


5-band image containing Cotton

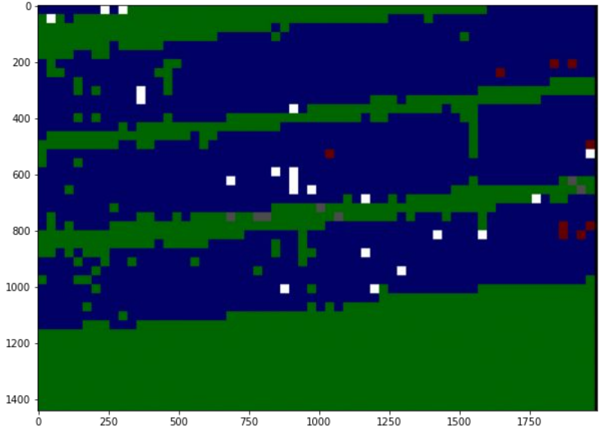
Comparison of CNN models on same images



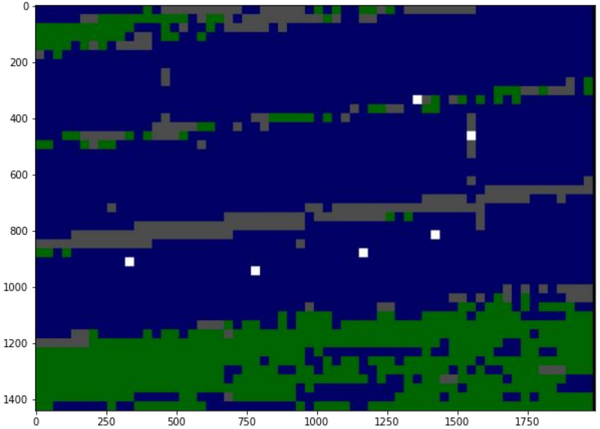
Original **Maize** Image



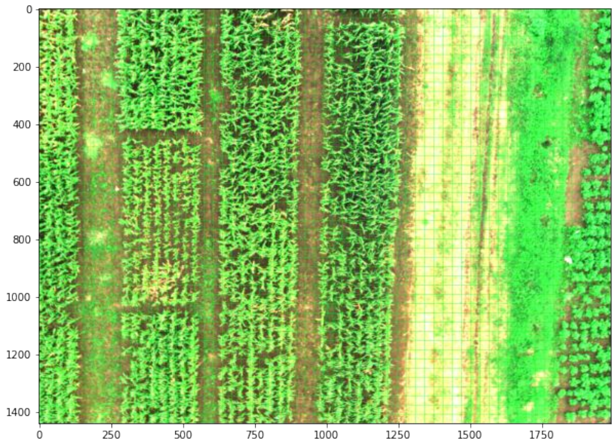
NIR CNN (model 1)



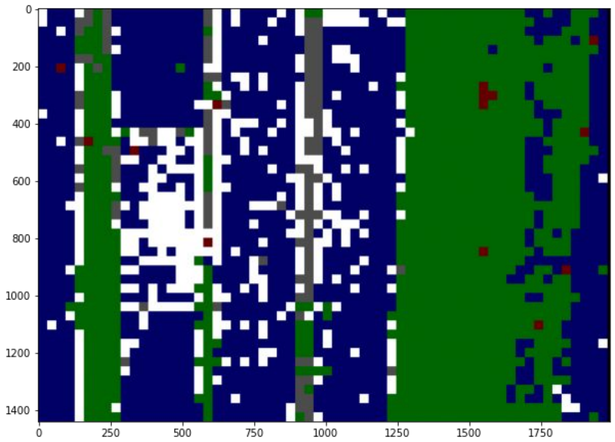
RGB CNN (Model 2)



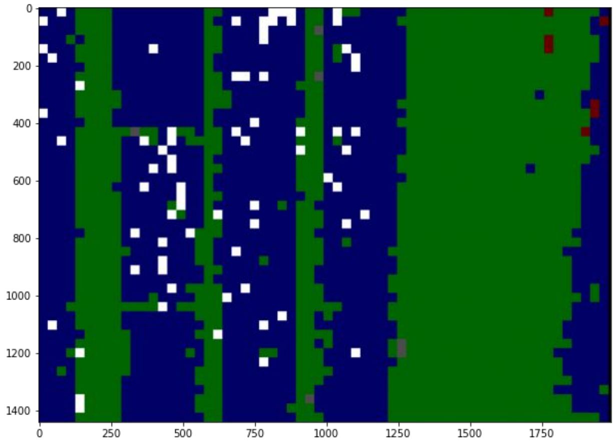
5-Band CNN (Model 3)



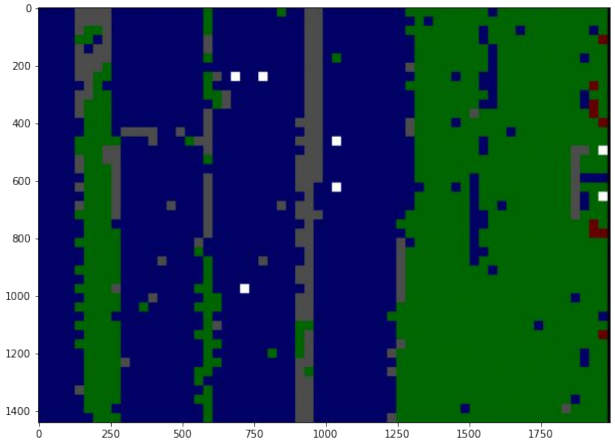
Original **Maize** Image



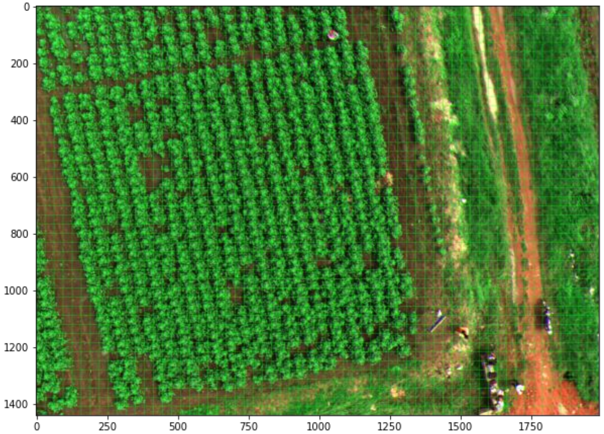
NIR CNN (model 1)



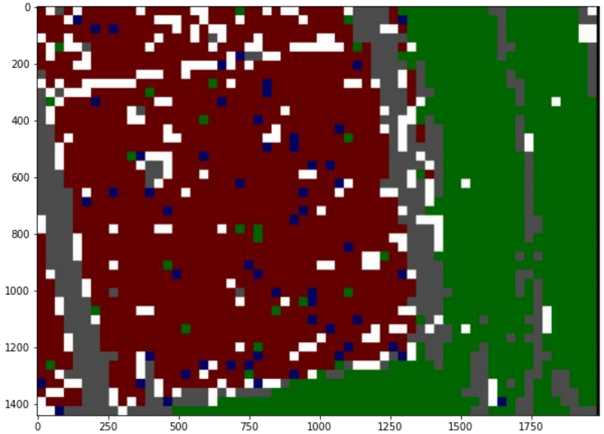
RGB CNN (Model 2)



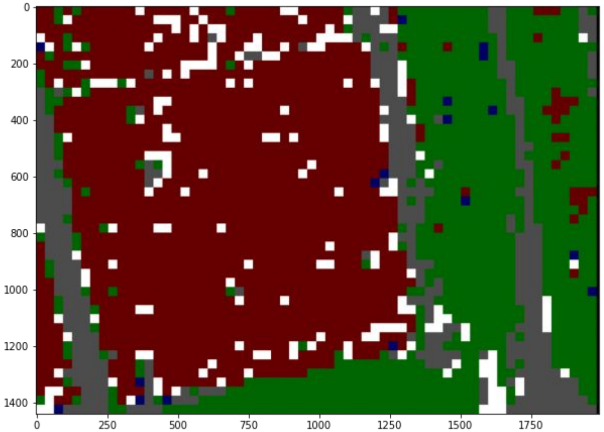
5-Band CNN (Model 3)



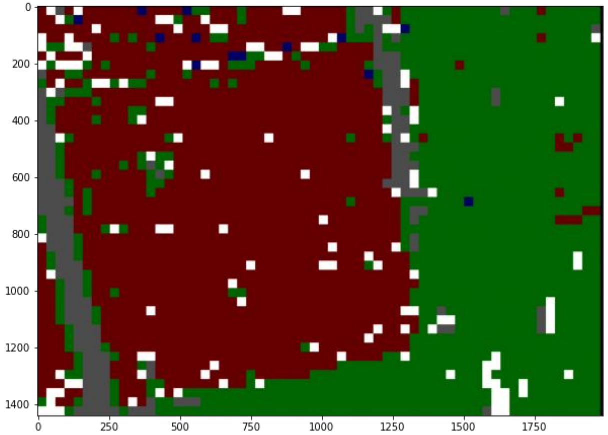
Original **Cotton** Image



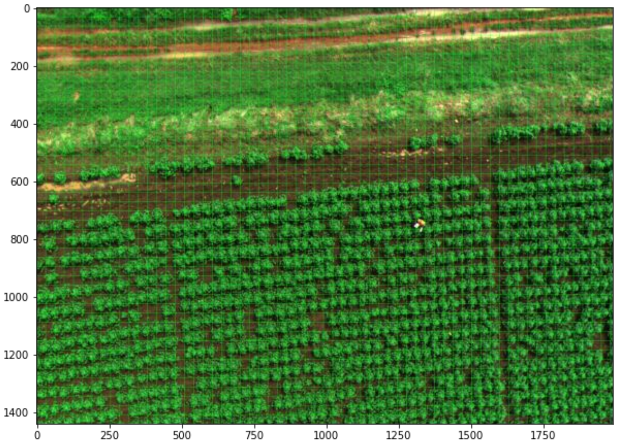
NIR CNN (model 1)



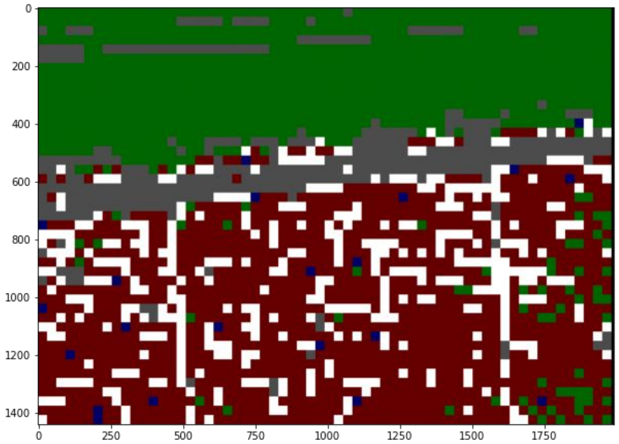
RGB CNN (Model 2)



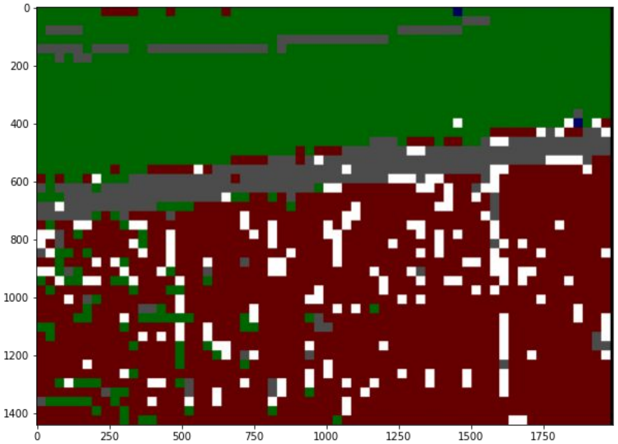
5-Band CNN (Model 3)



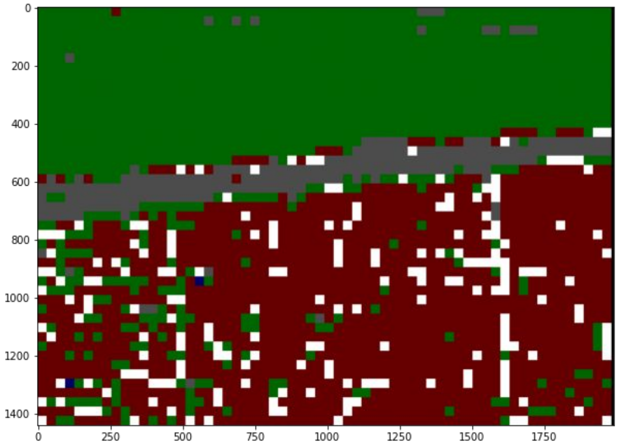
Original **Cotton** Image



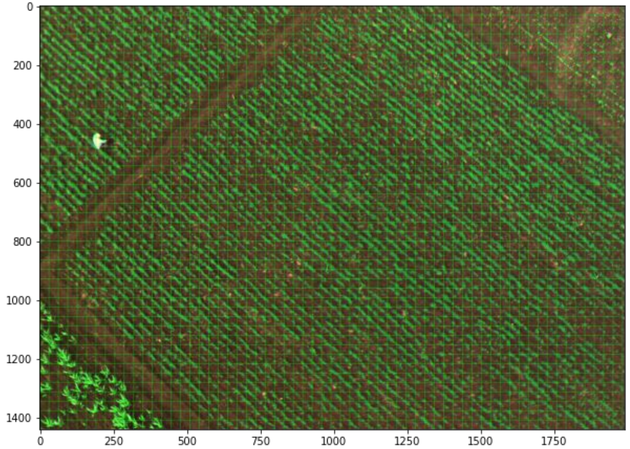
NIR CNN (model 1)



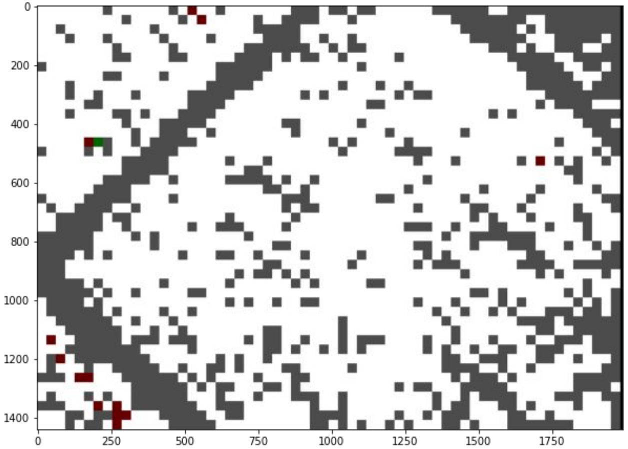
RGB CNN (Model 2)



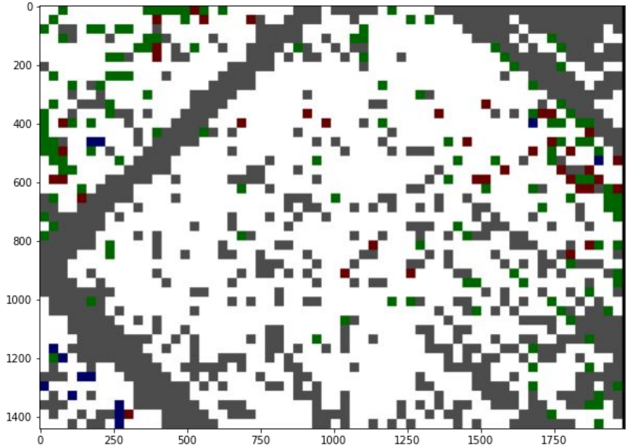
5-Band CNN (Model 3)



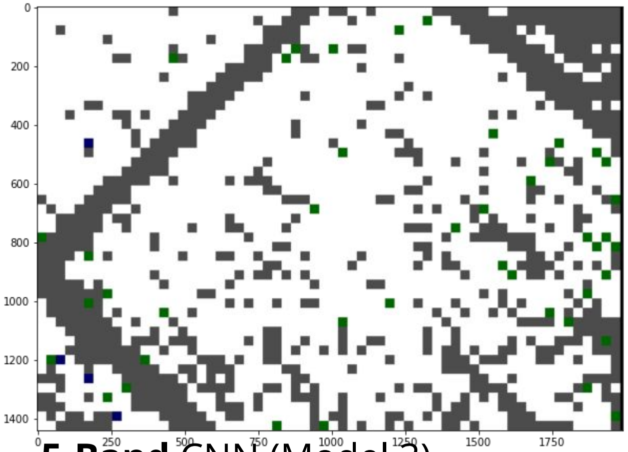
Original **Wheat** Image



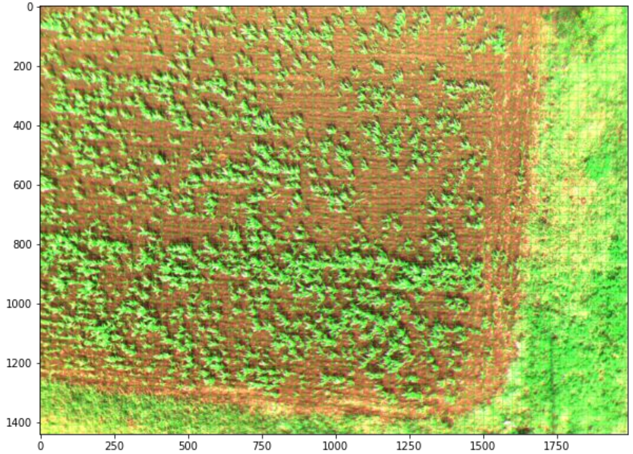
NIR CNN (model 1)



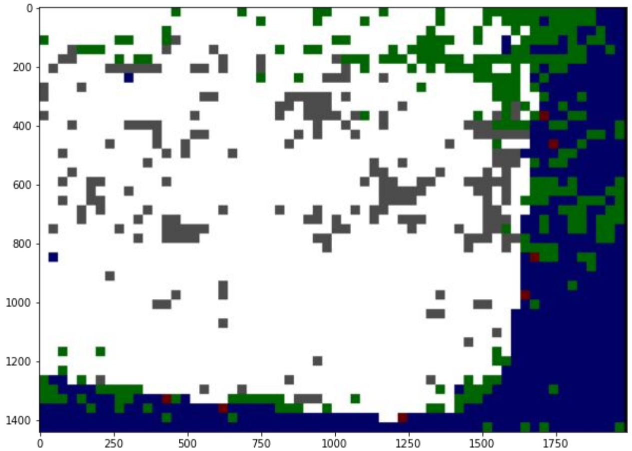
RGB CNN (Model 2)



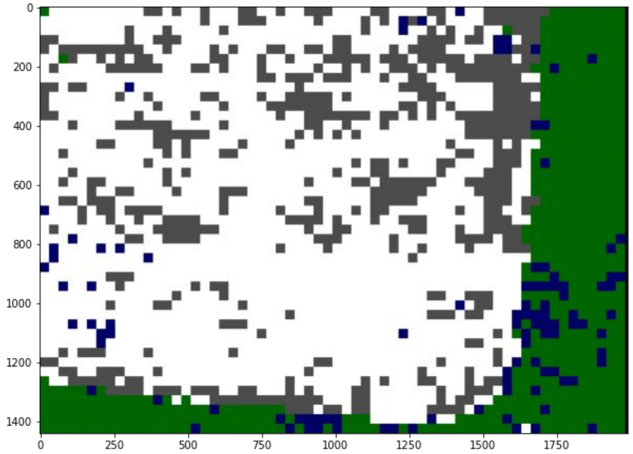
5-Band CNN (Model 3)



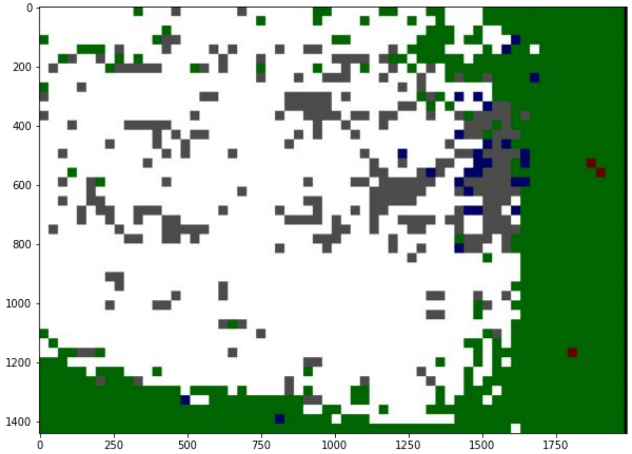
Original **Wheat** Image



NIR CNN (model 1)



RGB CNN (Model 2)



5-Band CNN (Model 3)

Wheat removal

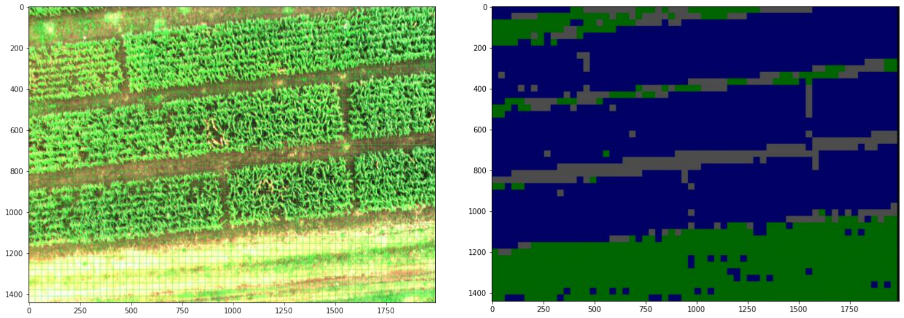
4.) CNN on 5-band images but without wheat class

The CNN models were getting confused with wheat and soil sometimes. So we tried removing wheat completely from the dataset and trying the model out. Following accuracies where attained-

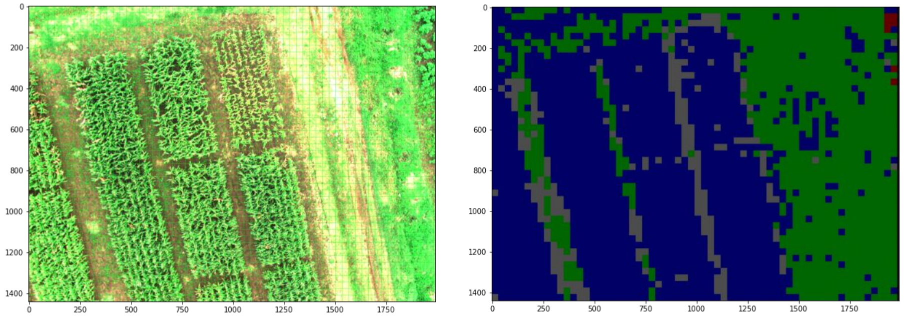
Train : 97.83%

Test : 96.94%

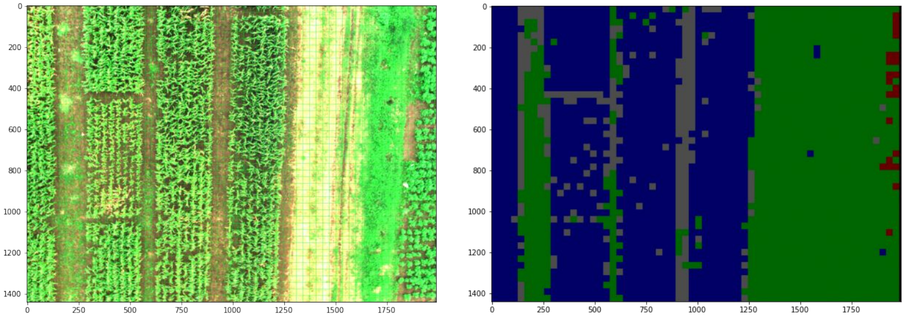
Model 4 (CNN without wheat)



5-Band image containing Maize

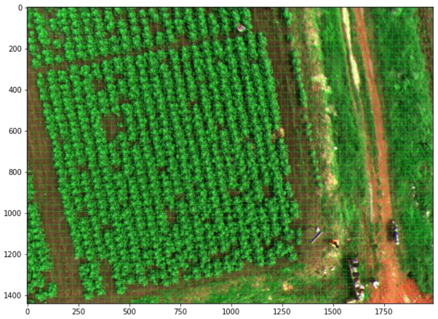


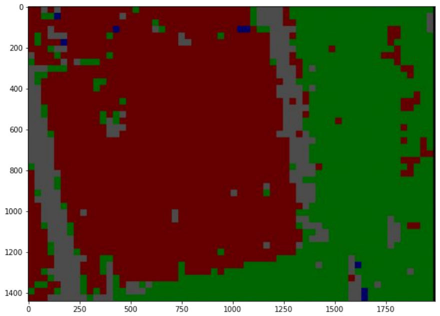
5-band image containing Maize



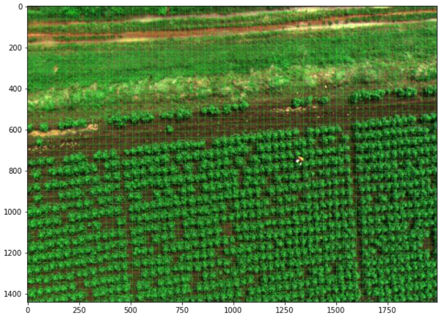
5-band image containing Maize

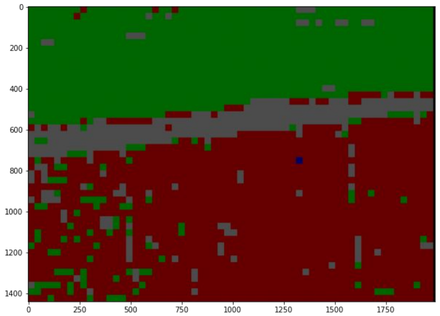
Model 4 (CNN without wheat)



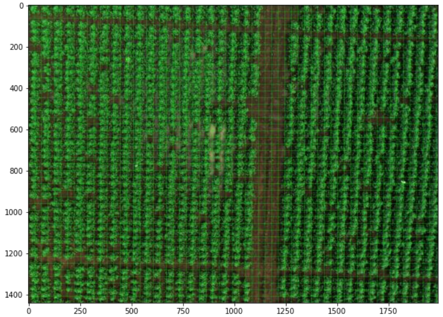


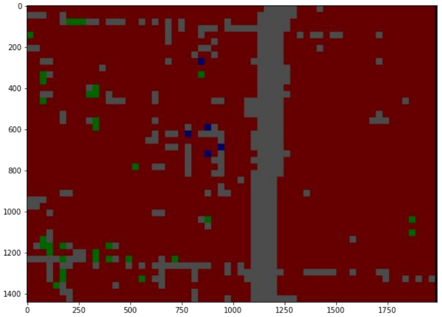
5-Band image containing Cotton





5-band image containing Cotton





5-band image containing Cotton

Random Forest

Random Forest

After CNN, we also experimented using a Machine Learning algorithm known as the Random Forest algorithm.

Random Forests operate by constructing a multitude of decision trees at training time and outputting the class that is the mode of the classes (classification) or mean/average prediction (regression) of the individual trees.

Random Forest

5.) RF on 5-band

An RF model was trained on all the 5-bands of the train images.
No. of estimators used: 100

Accuracies -

Train : 99.6%

Test : 91.8%

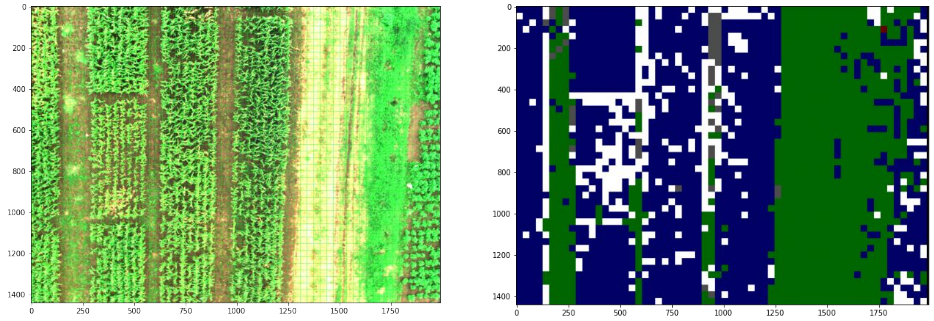
6.) RF on 5-band without Wheat images

The same model was trained on the dataset containing no wheat images .
No. of estimators used: 100

Accuracies -

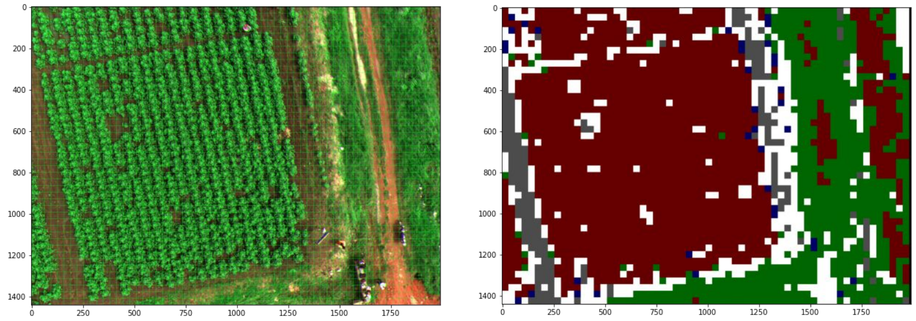
Train : 99.8%

Test : 94.3%

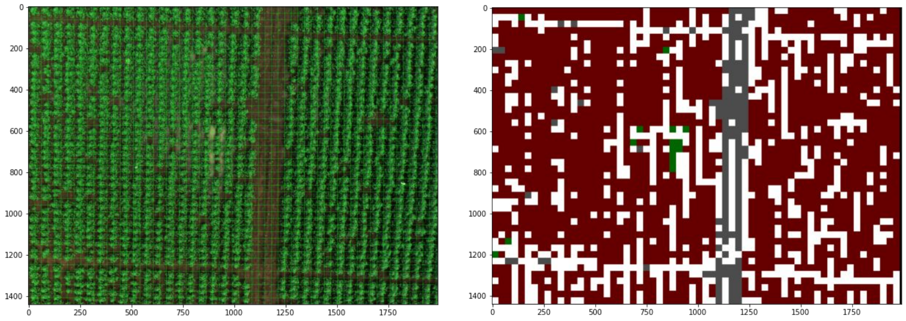


Model 5 (Random Forest with Wheat)

5-Band image containing Maize

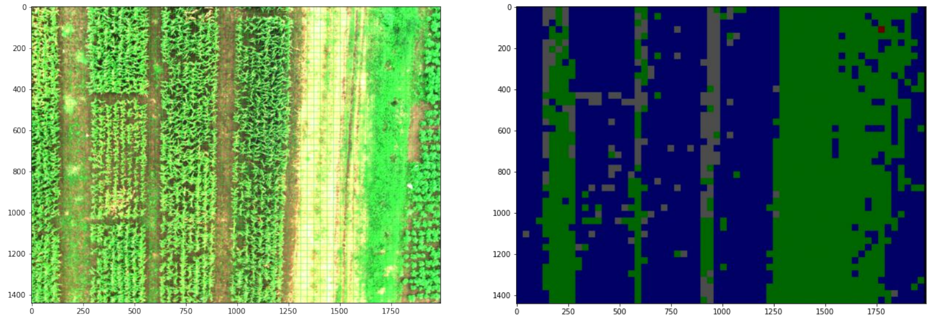


5-band image containing Wheat

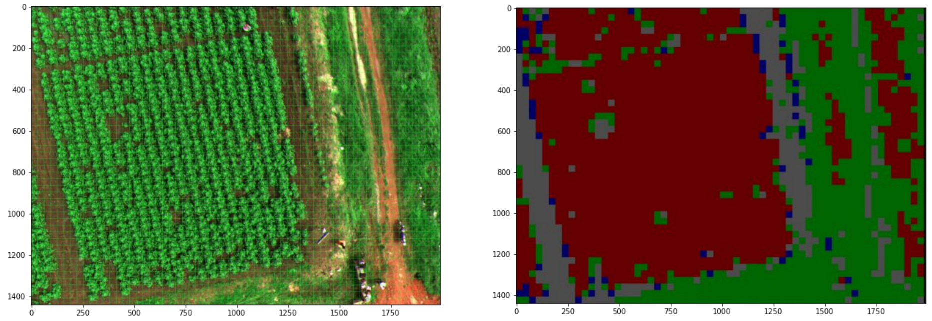


5-band image containing Cotton

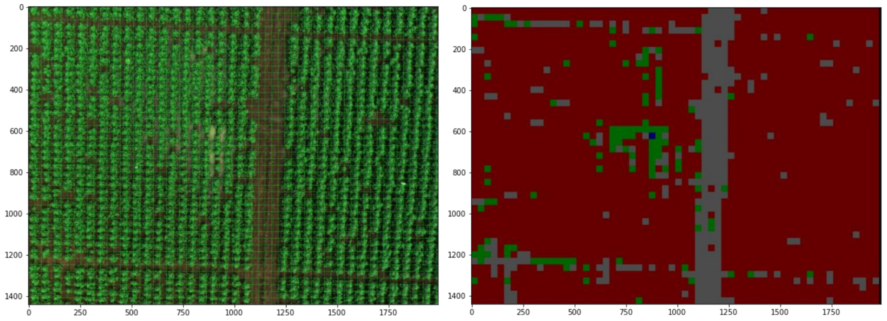
Model 6 (Random Forest without wheat data)



5-Band image containing Maize

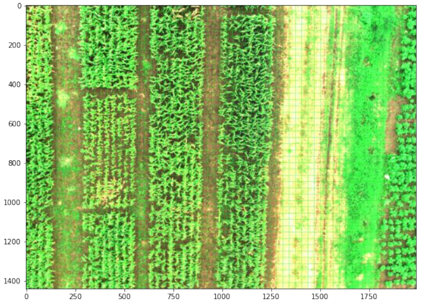


5-band image containing Cotton

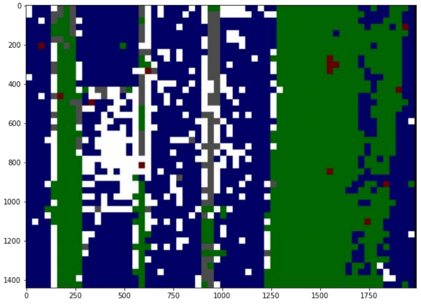


5-band image containing Cotton

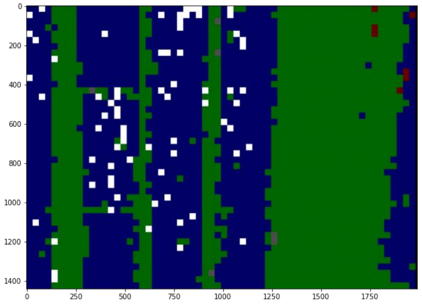
Comparison of all models



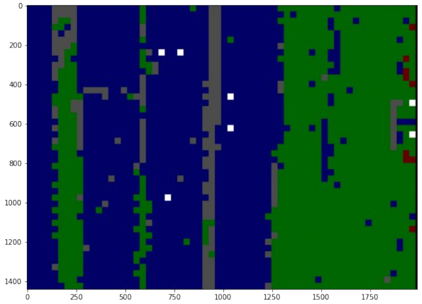
Original **Maize** Image



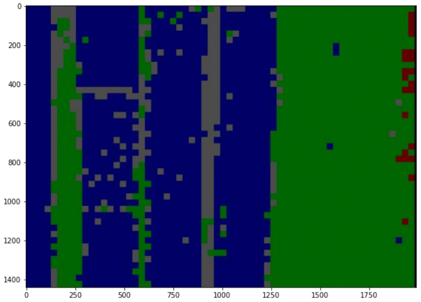
NIR CNN (Model 1)



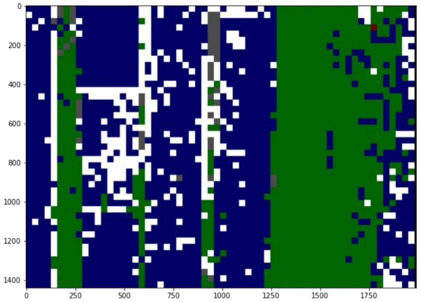
RGB CNN (Model 2)



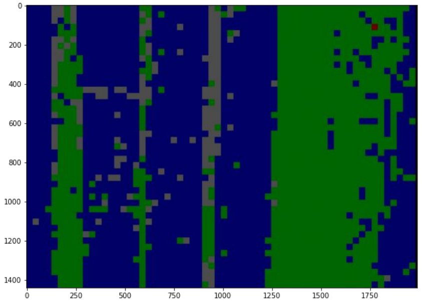
5-Band CNN (Model 3)



5- Band without wheat
CNN (Model 4)

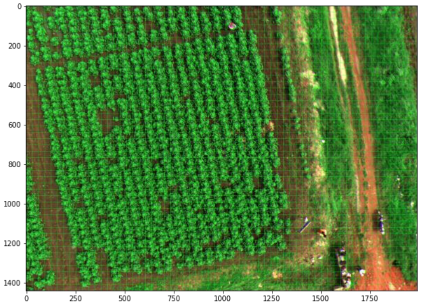


Random Forest
(Model 5)

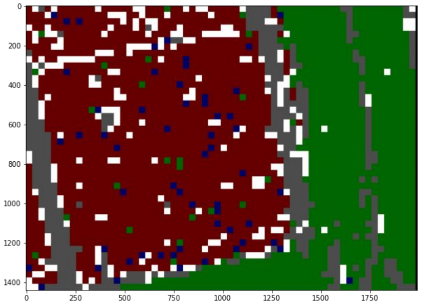


Random Forest without Wheat
(Model 6)

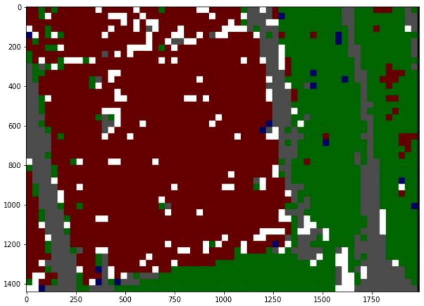
All the 6 models' output on a **Maize** image



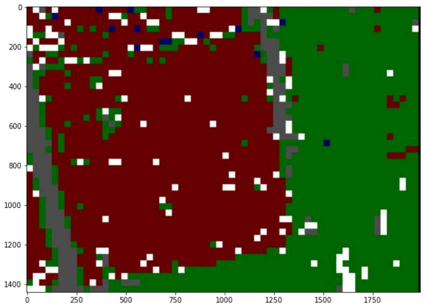
Original **Cotton** Image



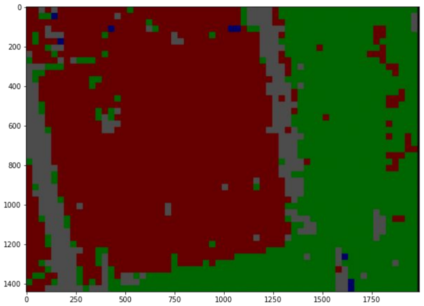
NIR CNN (Model 1)



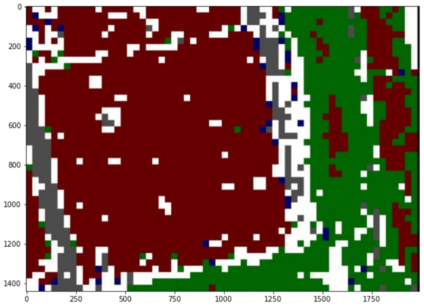
RGB CNN (Model 2)



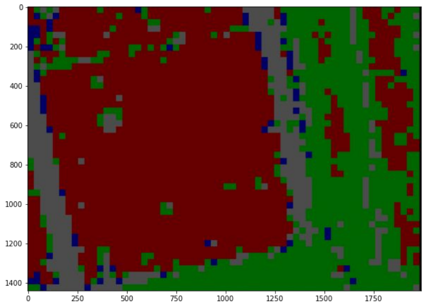
5-Band CNN (Model 3)



5- Band without wheat CNN (Model 4)

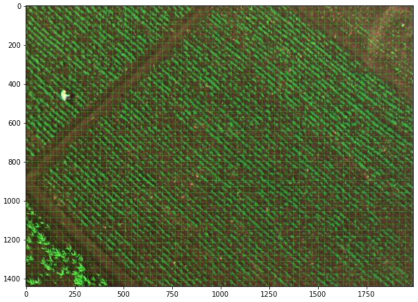


Random Forest (Model 5)

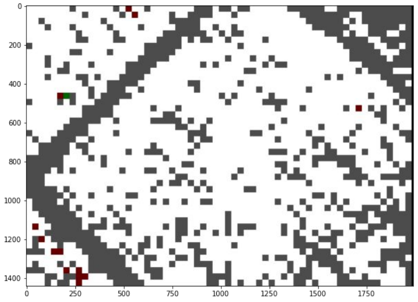


Random Forest without Wheat (Model 6)

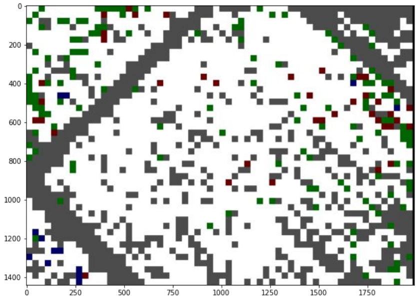
All the 6 models' output on a **Cotton** image



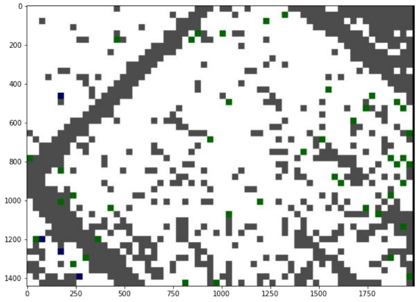
Original **Cotton** Image



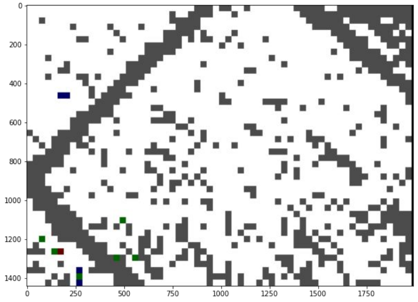
NIR CNN (Model 1)



RGB CNN (Model 2)



5-Band CNN (Model 3)



Random Forest
(Model 5)

Output of the 4 models which work on wheat, on a **Wheat** image

Model	Train accuracy	Test accuracy
CNN (NIR only)	95.3	90.3
CNN (RGB)	96.2	92.8
CNN (5 bands)	98.3	96.68
CNN (5 bands without Wheat)	97.83	96.94
RF (5 bands)	99.6	91.8
RF (5 bands without Wheat	99.8	94.3

Conclusion

Conclusions

- We experimented with different models and tuned various hyperparameters, and finally decided that SGD optimizer with a learning rate of 0.001 tended to work well. If the learning rate was higher, the testing accuracy and loss tended to oscillate.

- As our experiments show, using multispectral imaging helps in better classification of crop species. The 5-band models work better that RGB model or single band model .

Conclusions

- Before classification, the image has to be divided into around 2700 windows. The chosen classifier will be run on all of the 2700 windows, which might take some time. The Random Forest algorithm, which had about 100 estimators, took around 4-5 minutes to run on one large input image. Whereas CNN was faster in comparison and took about 1.5-2 minutes.

- We also constructed the NDVI map of an orthomosaic using the NIR band, which shows vegetative cover in the region.

Difficulties

- As mentioned before, Agisoft demo has its limitations and not a lot of good open source alternatives are present for mosaicing.

- Micasense provided all the functionality required for preprocessing the images. However, its installation was complicated and arduous. Moreover, the documentation is not too comprehensive.

- The sparse nature of wheat crops created problems while labelling. It confused the neural network to misclassify soil, an issue which vanished after removing wheat.

Further Work and Improvements

- Adding more Crop classes in the models

- Extending the classifiers to determine features other than crop species i.e Water/Nutrient levels, growth, yield of the cropping areas.

- Training a classifier which can take the entire field's Orthomosaic and produce output on it

Related work and references

1. A Photogrammetry Software as a Tool for Precision Agriculture: A Case Study
2. A Review on UAV-Based Applications for Precision Agriculture
3. Identifying Species and Monitoring Understorey from UAS-Derived Data: A Literature Review and Future Directions
4. Finer Classification of Crops by Fusing UAV Images and Sentinel-2A Data
5. Agisoft Orthomosaic tutorial
6. Micasense - GitHub

Thank you.