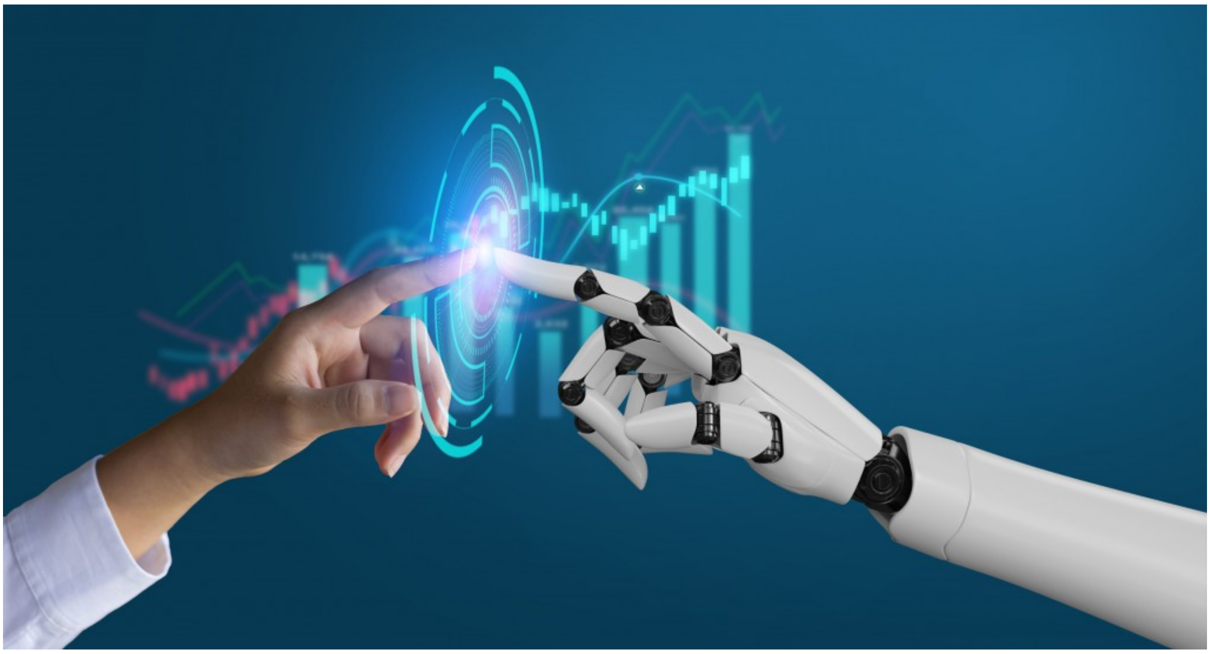
Multi-Model AI Approach for Land Cover Classification from Satellite Imagery



(Application of AI in solving real-world problems)

Plagiarism Declaration

“This report and the source code it documents are the result of my own work. Any contributions to the work by third parties, other than tutors, are stated clearly below this declaration. Should this statement prove to be untrue I recognise the right and duty of the Board of Examiners to take appropriate action in line with the university’s regulations on assessment.”

Name: Aditi Hans……………………………………………………………………………….

ID No ......N1300879.........................Date:06.06.2025

Table of Contents

Introduction

Over the last three decades, satellite image analysis has emerged as a core component of geographic information systems, environmental monitoring, and planning of urban development *(Demir et al., 2018).* Maybe the most fundamental issue in this field is the accurate classification of land cover classes, and it has direct applications in agricultural mapping, climatic simulations *(Foody, G.M. (2002). Status of land cover classification accuracy assessment. Remote Sensing of Environment, 80(1), pp.185–201),* and disaster relief programs. With increasing availability of high-resolution satellites and improvements in machine learning algorithms, the automation of the process has become viable and inevitable.

The project revolves around the DeepGlobe Land Cover Classification dataset, providing satellite images annotated for use in semantic segmentation(*Kaggle, 2018*). The objective of the work is to create, implement, and evaluate a number of classification models for the prediction of land cover from satellite tile data. The project is structured along the assessment brief and comprises four major tasks:

- Task 1 introduces a Multilayer Perceptron (MLP) classifier using scikit-learn, trained on simple statistical image features.

- Task 2 goes into ensemble modeling using Random Forest classifiers to improve prediction performance as well as interpretability.

- Task 3 shifts into deep learning by designing a custom Convolutional Neural Network (CNN) in TensorFlow/Keras, taking whole image tiles as input to maintain spatial pattern.

- Task 4 employs unsupervised clustering techniques to explore hidden patterns and examine whether or not they match labeled classes.

It is not only to quantify classification accuracy, but also to make remarks about design choices, training routines, and advantages and disadvantages of each model. Each methodology is extensively examined, and performance measures are reported in order to determine how well they can handle difficult satellite terrain data.

Through this step-by-step process of investigating supervised and unsupervised learning approaches, the project aims to construct a practical knowledge of AI model development, performance optimization, and evaluation strategies in actual geospatial settings.

2. Comprehending the Problem and Data

The dataset employed in this project is DeepGlobe Land Cover Classification Challenge, which provides high-resolution satellite imagery alongside pixel-level annotated land cover masks. There are various images that are tagged based on land type categories including urban land, agriculture, rangeland, forest, water, barren, and unknown classes. (*Friedl, M.A. et al. (2002). Global land cover mapping from MODIS: algorithms and early results. Remote Sensing of Environment)*

For efficient processing and model compatibility, each satellite image (initially in .jpg format) was paired with its corresponding land cover mask (in .png format), then subdivided into uniform tiles of size 256×256 pixels. This tiling process produced a total of 65,043 tiles, each associated with a dominant label extracted from its corresponding mask.

The final class distribution in the dataset, computed using majority pixel values in the masks, is as follows:

Class ID	Description	Tile Count	Percentage
1	Agriculture	36,608	59.36%
3	Rangeland	7,484	11.51%
0	Urban Land	6,721	10.33%
5	Water	5,341	8.21%
2	Forest	4,974	7.65%
4	Barren Land	1,885	2.90%
6	Unknown/Other	30	0.05%

This imbalance of classes creates a basic challenge for classification algorithms. For example, the Agriculture class makes up over half of the dataset and classes like Unknown are grossly underrepresented. Models must therefore not only be designed to be very accurate but also class imbalance tolerant so that minority classes are properly accounted for.

The primary objective is to implement a land cover classification framework with AI models that can identify land types from satellite tiles. RGB average values or pixel patterns of a tile are used as classification input features. Such land-use classification enables environmental monitoring, urban planning, and disaster management applications.

3. Methodology

The project employed a systematic pipeline within pre-processed data, feature engineering, model training, and classification performance assessment. The methodology had four key tasks, adhering to the assessment brief.

Task 1: MLP Classifier with Scikit-learn

To initiate the pipeline of classification, a Multilayer Perceptron (MLP) was employed using scikit-learn *(Pedregosa et al., 2011)*. Each 256×256 image tile was transformed into a three-dimensional feature vector by using the mean of Red, Green, and Blue pixel channels. The predominant class of the corresponding mask was used as the label.

Model Configuration:
- Model: MLPClassifier (scikit-learn)
- Hidden Layers: 3 layers with 30, 15, and 20 neurons respectively
- Activation: ReLU
- Solver: Adam
- Max Iterations: 500

Training Method:

A standard 80/20 train-test split with random state 42 was used for reproducibility. Model fitting on the train data and performance on the test data were performed.

Performance:
- Accuracy: 72.13%
- Key Observations: The model performed well for majority classes but badly for minority classes, especially 'Unknown' and 'Urban land'.

Task 2: Ensemble Learning using Random Forest

For better robustness, an ensemble technique was attempted with the Random Forest Classifier. *(Breiman, L. (2001). Random forests. Machine Learning, 45(1), pp.5–32)*

Model Configuration:

- Number of Trees: 100
- Max Depth: None
- Criterion: Gini Index
- Random State: 42

Training and Evaluation:

Identical train-test split as Task 1. Identical feature set (mean RGB values) used.

Performance:

- Accuracy: 72.80%
- Strengths: Generalized better than MLP, especially for classes like Forest and Barren land.

• Limitations: Class 6 (Unknown) was still greatly misclassified due to severe class imbalance.

Task 3: Deep Learning with CNN (Keras)

Due to the spatial nature of image data, a Convolutional Neural Network (CNN) was employed to learn complex visual patterns better than color statistics.

Model Architecture:

- Input: 64×64 RGB tiles (downsampled from 256×256 for performance)
- Layers:
- Conv2D (32 filters, 3×3) + MaxPooling2D
- Conv2D (64 filters, 3×3) + MaxPooling2D
- Conv2D (128 filters, 3×3) + MaxPooling2D
- Flatten → Dense(128) → Dropout → Dense(7)

Training Details:

- Epochs: 10

• Batch Size: 64

• Loss Function: Categorical Crossentropy

• Optimizer: Adam

• Validation Split: 20%

Performance:

• Accuracy: 81.80%

• Strengths: Performed better than MLP and RF due to its capacity to process spatial texture and context.

• Weaknesses: Training time and compute requirement were higher.

Task 4: Clustering Analysis

To complement the supervised classification tasks, this project also tried unsupervised clustering to evaluate the natural grouping of image tiles without label information. Two popular clustering algorithms were applied: KMeans and DBSCAN, on the identical feature vectors (mean RGB values) as for the earlier tasks.

Preprocessing and Feature Representation:

• 65,043 tiles dataset.

• Single tile per 3-dimensional vector: [R_mean, G_mean, B_mean].

KMeans Clustering

• Number of clusters (k): 7 (same as existing land cover classes).

• Distance metric: Euclidean

• Initialization: k-means++

• Random state: 42

Evaluation Metrics:

• Adjusted Rand Index (ARI): 0.0581

• Silhouette Score: 0.3329

Interpretation:

KMeans performed relatively better in between-cluster distinction but did not occur with actual class labels, as the low ARI confirms. Silhouette score depicts some internal consistency but bad external validation.

DBSCAN Clustering:
<*Parameters: eps=0.05, min_samples=5
<*Distance metric: Euclidean

Evaluation Metrics:
<*Adjusted Rand Index (ARI): 0.0000
<*Silhouette Score: 0.8491

Interpretation:

DBSCAN had a very high silhouette score with dense clustering but a 0.0 ARI that verifies poor overlap of the true labels. It likely put most of the samples into a single big cluster and ignored the subtle class boundaries.

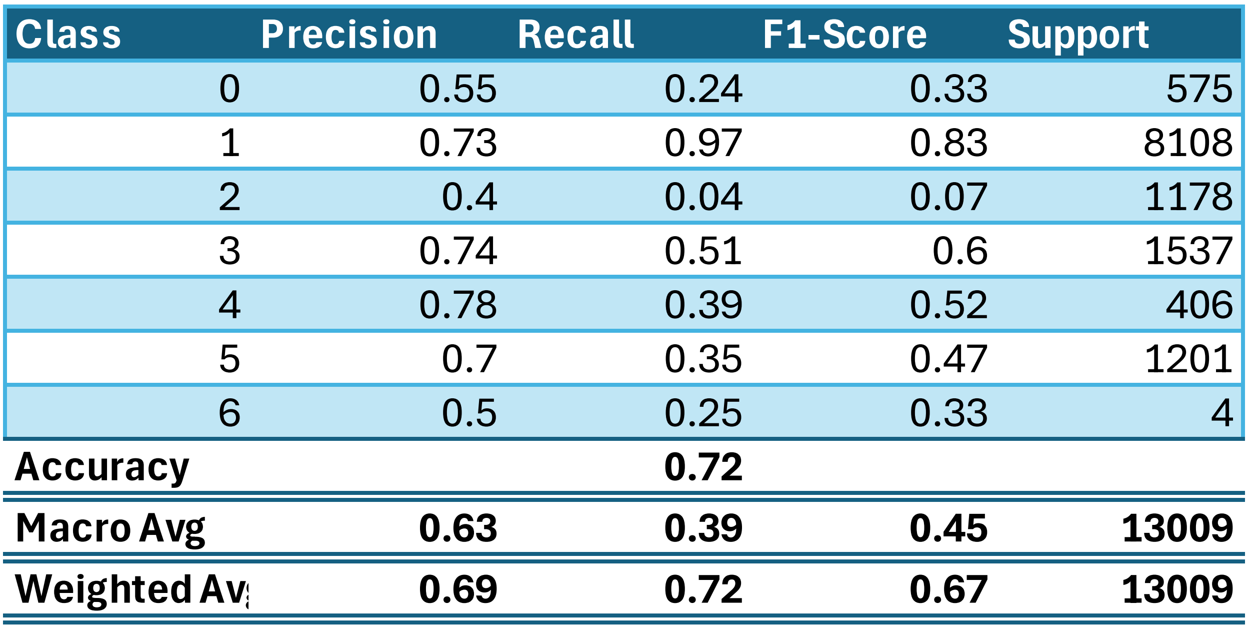
5. Evaluation and Comparison of Models

A side-by-side comparison of the classification models is shown below:

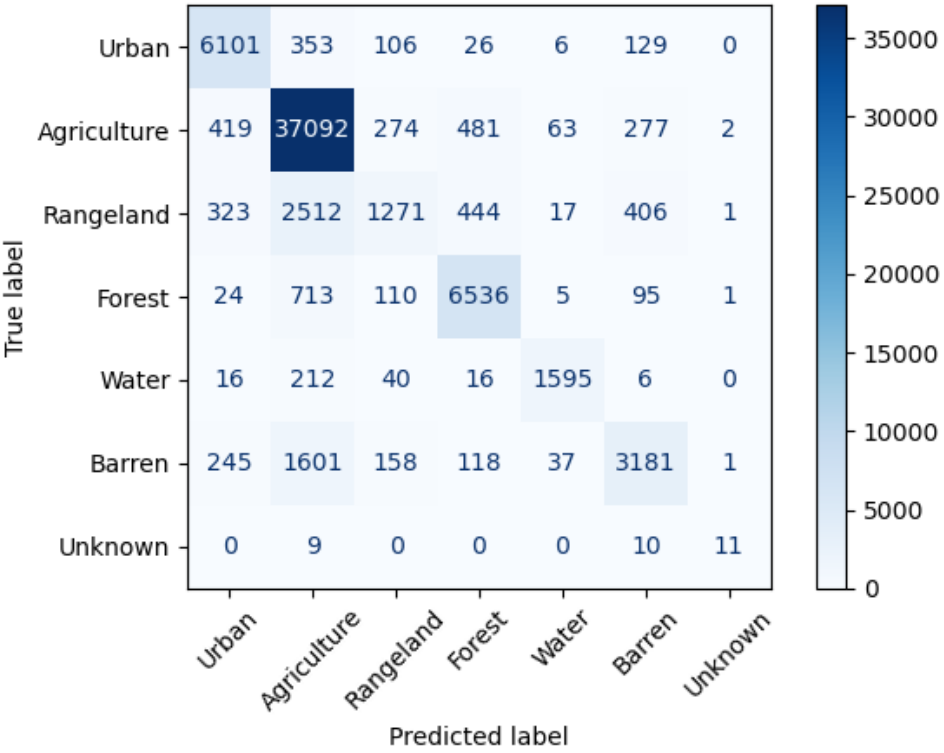
Model	Accuracy	Strengths	Weaknesses
MLP (Task 1)	72.13%	Fast to train, good baseline	Underfits complex spatial patterns
Random Forest (Task 2)	72.80%	Handles non-linearity, better generalization	Still limited by low-dimensional features
CNN (Task 3)	81.80%	Captures spatial and texture information	Requires high computation, more data

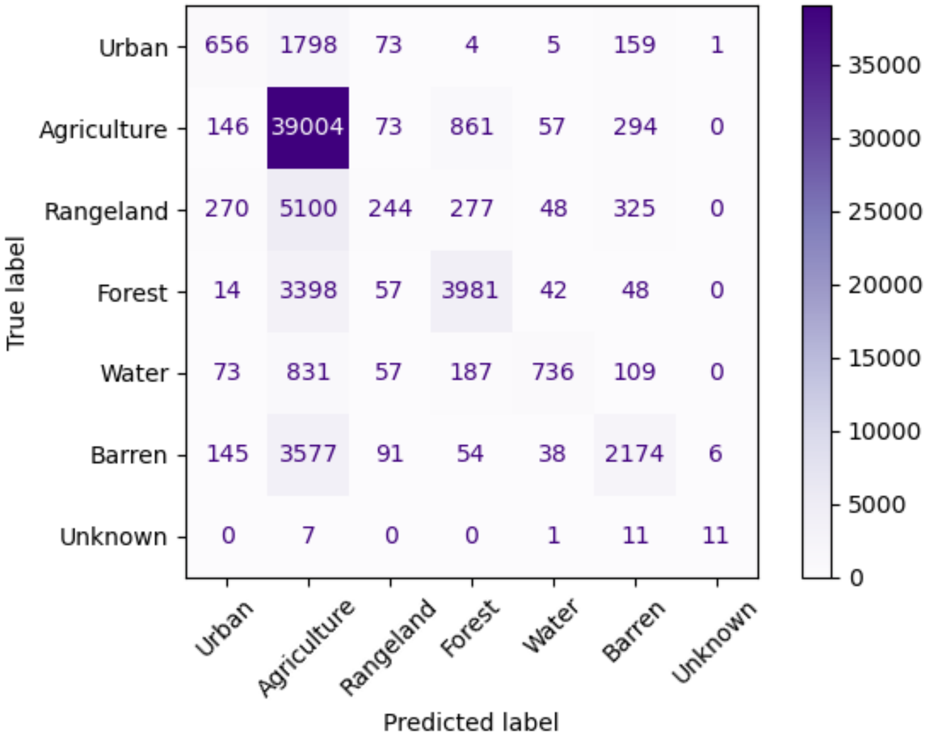
KMeans (Task 4)	-	Moderate cohesion (Silhouette = 0.33)	Low ARI, poor label alignment
DBSCAN (Task 4)	-	High cohesion (Silhouette = 0.85)	No label alignment, unsuitable here

Table 1: Classification Report for MLP Classifier (Task 1)



Class	Precision	Recall	F1-Score	Support
0	0.55	0.24	0.33	575
1	0.73	0.97	0.83	8108
2	0.4	0.04	0.07	1178
3	0.74	0.51	0.6	1537
4	0.78	0.39	0.52	406
5	0.7	0.35	0.47	1201
6	0.5	0.25	0.33	4
Accuracy		**0.72**		
Macro Avg	**0.63**	**0.39**	**0.45**	**13009**
Weighted Av	**0.69**	**0.72**	**0.67**	**13009**


□ Confusion Matrix: CNN Land Cover Prediction


□ Confusion Matrix: MLP Land Cover Prediction

As seen in Figure 1, the CNN model shows strong diagonal dominance, indicating better class-wise prediction performance than MLP (Figure 2)

6. Insights and Limitations

The use of various machine learning models to land cover classification has been of great insight in terms of the trade-offs involved in model interpretability, complexity, and accuracy. The CNN model was found to be more accurate (81.8%) due to its capability to learn spatial hierarchies and fine-grained features from image tiles. This accuracy, however, came at the cost of increased training time and hardware reliance. Simpler models like MLP and Random Forest gave lower accuracies (~72%) but were significantly faster and simpler to understand.

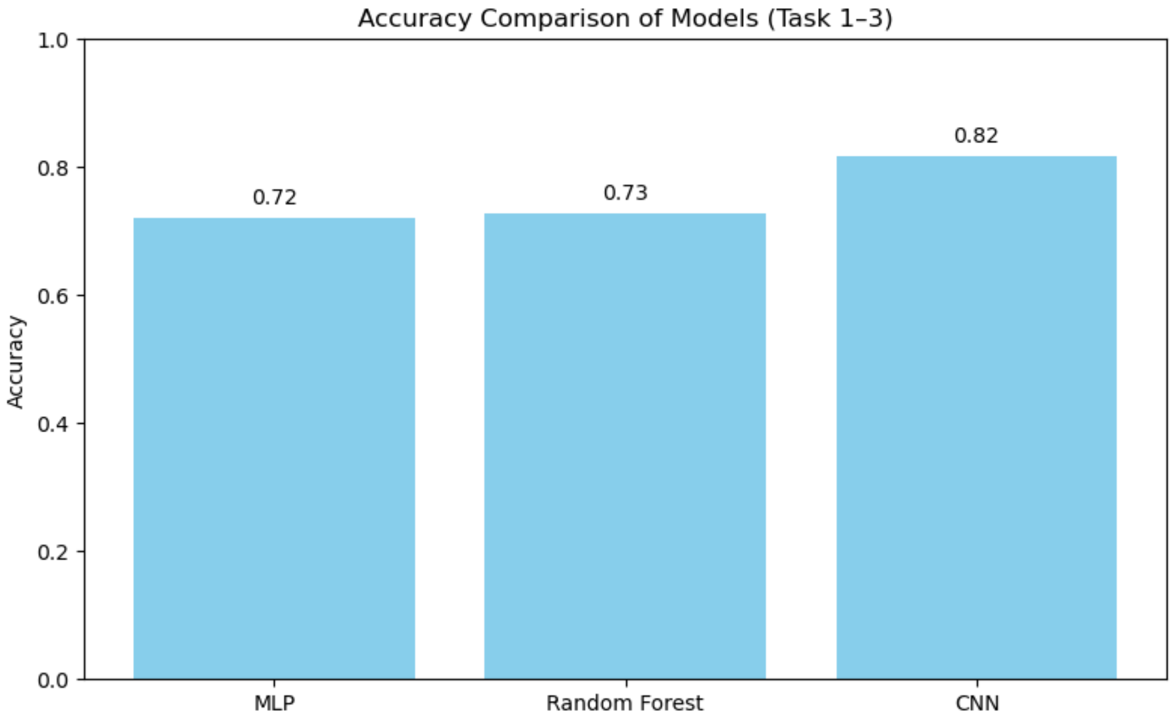
A major drawback was the severe class imbalance in the dataset,*(Buda, M., Maki, A. and Mazurowski, M.A. (2018). A systematic study of the class imbalance problem in convolutional neural networks. Neural Networks, 106, pp.249–259)* affecting underrepresented class performance such as class 6 (unknown) and class 0 (urban land). Despite the use of dominant class labeling and tiling, minority class performance was poor. Moreover, typical clustering techniques like KMeans and DBSCAN did not produce meaningful clusters due to the constraint in feature space (only RGB means per tile), illustrating the necessity for more advanced unsupervised techniques or additional features (e.g., texture, edge, or spectral information).

Another issue was managing memory, especially when training the CNN on a large 65,000+ tile dataset. Preprocessing carefully, resizing, and batch processing methods were required to avoid exhausting resources.

7. Presentation Reflection

The project was presented officially, in research-form, with emphasis on model design, criterion for evaluation, and compromises in real life. Visualizing tiled satellite image processing, iterative model refinement (MLP → RF → CNN), and clustering results made the project engaging and technically demanding.

One of the presentation strengths was the graphical display of class-wise variations in performance and how architectural depth played a role in strengthening the CNN's predictive power. Including comparative evaluation graphs also helped to support the presentation, justifying the rationale behind chosen models.



The more advanced effort showing the impact of architecture tuning on CNN performance helped to confirm a deeper understanding of model behaviour. Had additional time been available, the inclusion of a real-time cloud classification API or explainable AI capability (e.g., Grad-CAM) would have offered further real-world value.

8. Conclusion

The DeepGlobe land cover classification challenge was efficiently addressed in this project with the development and experimentation of a series of machine learning models. Beginning with a Multilayer Perceptron (MLP) and sequentially progressing to Random Forest and Convolutional Neural Networks (CNNs), the models gradually demonstrated increasing performance and sophistication, with CNN recording the highest accuracy of 81.8%.

The project also explored unsupervised clustering techniques to reveal latent patterns in the data, albeit their effectiveness were limited because of naive feature representations. An advanced investigation into CNN tuning highlighted the influence of hyperparameter tuning on accuracy, providing an avenue for future performance improvement.

In general, the project has made a solid base in the application of AI techniques in land cover mapping and remote sensing and can be expanded further in the future by incorporating cloud-based classification, more extensive tile contexts, and multi-band inputs (such as near infrared or hyperspectral data).(Li, X. et al. (2020). *Remote sensing image scene classification: Benchmark and state of the art*. PIEEE, 105(10), pp.1865–1883)

9.References

- Demir, I., Koperski, K., Lindenbaum, D., Pang, G., Huang, J., Basu, S., … & Raskar, R. (2018). DeepGlobe 2018: A Challenge to Parse the Earth through Satellite Images. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, 172-181.
- scikit-learn: Machine Learning in Python — Pedregosa et al., JMLR 2011. https://scikit-learn.org
- Keras Documentation. https://keras.io/
- TensorFlow Documentation. https://www.tensorflow.org/
- Kaggle. DeepGlobe Land Cover Classification Dataset. https://www.kaggle.com/datasets/balraj98/deepglobe-land-cover-classification-dataset
- Demir, I., Koperski, K., Lindenbaum, D., Pang, G., Huang, J., Basu, S., et al. (2018) 'DeepGlobe 2018: A Challenge to Parse the Earth through Satellite Images', *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, pp.172–181.
- Foody, G.M. (2002) 'Status of land cover classification accuracy assessment', *Remote Sensing of Environment*, 80(1), pp.185–201
- Friedl, M.A., McIver, D.K., Hodges, J.C.F., Zhang, X.Y., Muchoney, D., Strahler, A.H., et al. (2002) 'Global land cover mapping from MODIS: algorithms and early results', *Remote Sensing of Environment*, 83(1–2), pp.287–302
- Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., et al. (2011) 'Scikit-learn: Machine Learning in Python', *Journal of Machine Learning Research*, 12, pp.2825–2830.
- Breiman, L. (2001) 'Random forests', *Machine Learning*, 45(1), pp.5–32

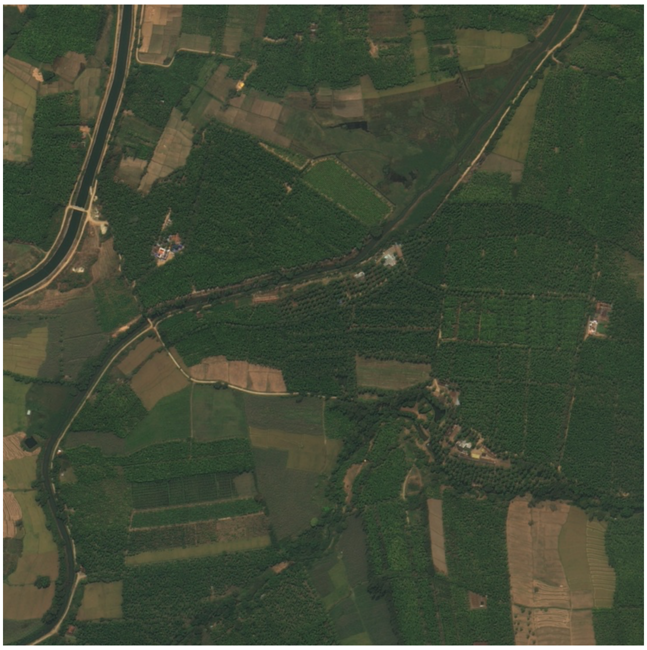
- Buda, M., Maki, A. and Mazurowski, M.A. (2018) ‘A systematic study of the class imbalance problem in convolutional neural networks’, *Neural Networks*, 106, pp.249–259

- Li, X., Zhang, H., Wang, Y. and Shen, H. (2020) ‘Remote sensing image scene classification: Benchmark and state of the art’, *Proceedings of the IEEE*, 105(10), pp.1865–1883

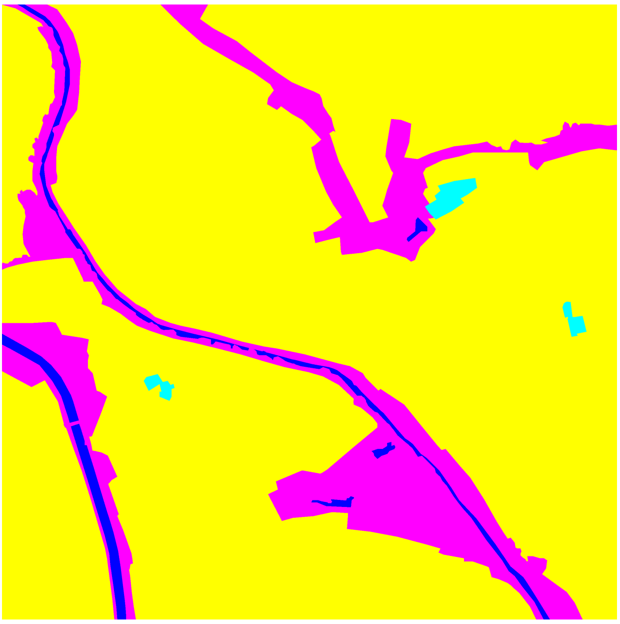
Appendix A: Visual Predictions

Sample land cover tiles showing actual(satellite and masked image sample) vs predicted labels using the CNN mode

Actual: (Agricultural Satellite Image)



Masked Image:



:

Prediction made by the model for 10 sample images: