



Land Use Land Cover Mapping using UAS Imagery: Scene Classification and Semantic Segmentation

A DIRECTED INDEPENDENT STUDY

Sudhagar Nagarajan

Civil, Environmental & Geomatics Dept
Florida Atlantic University
Boca Raton, FL 33486, USA
snagarajan@fau.edu

Monica Rajkumar

Civil, Environmental & Geomatics Dept
Florida Atlantic University
Boca Raton, FL 33486, USA
mrajkumar2020@fau.edu

ABSTRACT OF THE RESEARCH PROJECT

To determine the state of surface of earth, LULC classification plays a vital role. Vegetation Classification can be performed by incorporating various deep learning models using Convolution Neural Network approach. The primary purpose of this research is to find out the impressive performances of pre-existing deep Convolution Neural Network (CNN) models for vegetation classification in the Jupiter Inlet Lighthouse Outstanding Natural Area (JILONA). Specifically, this study focuses on examining of the capacity of Scene and Pixel classification technique using certain 3 band and 5 band combinations. Eight well known scene-based deep convnets, namely VGG19, ResNet152V2, InceptionV3, EfficientNetB5, Xception, InceptionResNetV2, MobileNetV2, DenseNet201, and one important pixel classification model, UNet for vegetation mapping in the North and South part of the JILONA. Among all the scene classification models, EfficientNetB5 and DenseNet201 outperformed with an accuracy of over 97%, Xception and ResNet152V2 model with 95%, whereas the remaining models ranging from 85% to 92%. The classification map achieved through UNet pixel-based method, resulted with an accuracy of about 87% and 91% when all 5 bands (Blue, Green, Red, RedEdge, Near infrared) were used for training.

Keywords Machine Learning, Deep Learning, CNN, UAS Imagery, LULC Classification.

1 Introduction

Jupiter Inlet Lighthouse Outstanding Natural Area (JILONA) is a region, which is covered by various kinds of vegetation, wildlife and other cultural components that rely on adequate land cover. Knowing the importance of LULC classification, there are many effective methods to carry out the vegetation mapping, namely satellite-based, aircraft-based, terrestrial-based etc., Recently, Deep learning, an asset of machine learning, is considered to be the spotlight of remote sensing field. This is because CNN had made a series of breakthroughs in various remote sensing applications such as classification, object detection and segmentation with its superior performance. The objective of this project is to demonstrate the capability of LULC classification by various deep learning models using UAV Imagery and determine the best CNN model to perform such vegetation classification. This is important when a particular site needs to be examined for improvements and in preserving the condition of the land. The main goal of this research is therefore examining the power of deep CNN for the land use and land cover classification based on UAV Imagery, investigating the generalization capacity of the existing CNN models and compare the efficiency of certain well known deep CNN scene-based models, which includes, VGG19, ResNet152V2, InceptionV3, InceptionResNetV2, MobileNetV2, DenseNet201, Xception, EfficientNetB5 and perform a study on UNet, known to be a efficacious Pixel based classification model. Thus, this study results in mapping the vegetation cover by a comprehensive and elaborative analysis using Unmanned Aerial remote sensing data acquired during the years 2018 and 2020.

2 Material and Methods

The data used for this study is the UAS Imagery data collected from Micasense RedEdge camera. The primary purpose of this research is to use this imagery and classify the land cover classes. Being a subset of machine learning, Deep learning deals with various algorithms that includes neural networks, which was inspired by the structure and function of a human brain. In all the data science technique including Deep learning algorithms performance will be based on the amount of data. In other words, accuracy improves when the training data is higher. Deep learning is often referred to as “Feature learning”, which is the ability of extracting features from raw data automatically. Whenever the inputs and output are analog, deep learning excels in its precision. The analog data includes imagery with pixels, text data, audio files etc., This is because, the model gains knowledge from the input data through a general-purpose learning procedure, without the intervention of human engineers.

One of the deep learning packages and a high-level API for tensorflow, “Keras applications”, consists of various models that have pre-trained weights which can be further used for different processes including feature extraction and prediction. Certain other python modules such as pandas, scikit-learn, matplotlib, seaborn, pickle etc., were all used for this study. The aim of this research is to investigate the efficiency of some of the Keras deep learning models in vegetation classification along the JILONA region, which will be a scene-based classification approach. Apart from this, a pixel-based classification approach is also performed for this study to come up with a classification map. This approach showcases and compares different combination of bands available from the Micasense RedEdge camera. The available band in the sensor is clearly shown in Figure 1. From these set of band information, three major combinations are identified, and the accuracy of the trained models are investigated. The selected band combinations are (1) Red, Green, Blue; (2) Green, Red, Near Infrared; (3) Blue, Green, Red, RedEdge, Near Infrared.

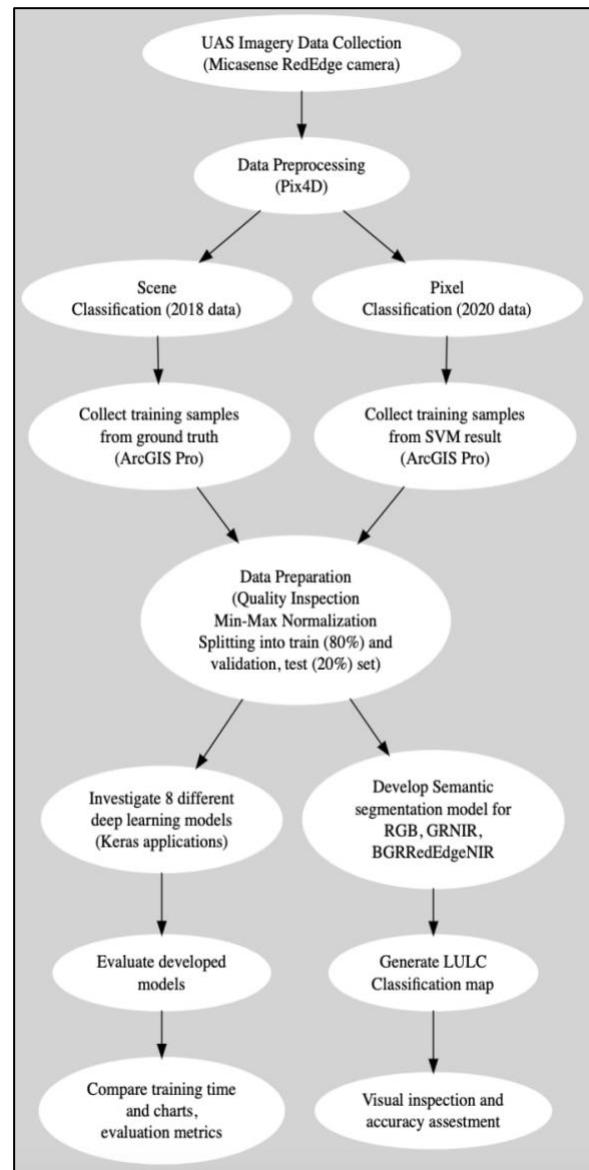


Figure 1: Research Methodology / Workflow

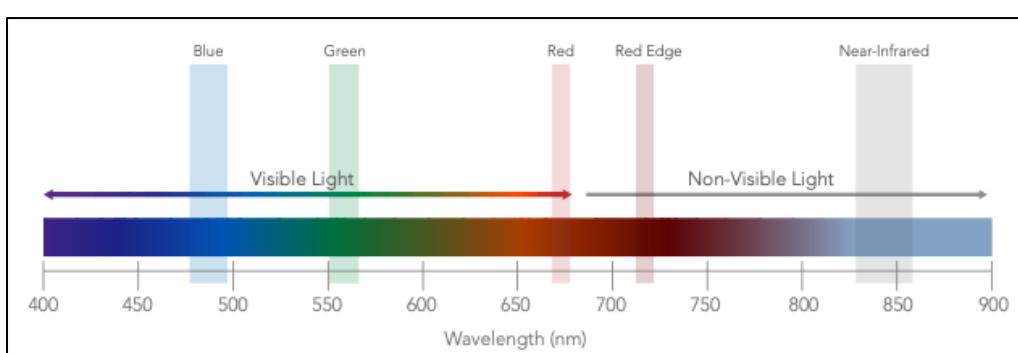


Figure 2: Visible to Non-visible Light Spectrum Range Represented (400 nm - 900 nm)

Initially, the raw data from the Micasense RedEdge camera was processed using Pix4D, in order to generate an orthomosaic imagery. This was followed by the process of collecting training samples for both scene-based and pixel-based classification methods. Training samples were collected separately based on the requirement of model. Thereafter, the collected samples were utilized to train the deep learning models, so that the accuracy of the models can be compared. While training, the process of augmentation was carried out in order to increase the training samples. Adam optimizer, softmax activation function, categorical cross entropy loss functions were all the major hyperparameters that were examined during the model training procedures.

In addition to this, for both scene-based and pixel-based classification method, 80% of the total dataset was considered to be the training set whereas the remaining 20% was allocated to validation and testing sets equally. Moreover, the acquired ground truth data was also considered to test the accuracy of the developed pixel-based classification model. The evaluation metrics such as accuracy, intersection over union (iou) and dice co-efficient were analyzed while testing the model.

The output from the scene-based classification model would be a name of a class that the input imagery belongs to, based on the probability, whereas for pixel-based classification model, the result would be the class name for each pixel. Hence, using pixel-based method, the classification map of whole region can be determined easily and accurately.

2.1 Scene Based Classification

Scene classifier is a deep learning technique in which scenes from the collected imagery data are classified categorically. When Object based method classifies prominent objects in foreground, scene classification utilizes the layout of available objects within the given image along with the ambient context. It is often referred to as scene analysis or scene recognition.

2.1.1 VGG19

VGG19 is one of the variants of Visual Geometry Group (VGG) model. VGG19 model comprises of 19 layers, which includes 16 convolution layers, 3 Fully connected layer, 5 MaxPool layers and 1 SoftMax layer. Similar kinds of variants include, VGG11, VGG16 etc., It results in improved accuracy when deep convolution neural layers are used. Spatial padding, Max pooling and convolving were all used in training this model, which was followed by Rectified linear unit (ReLu) activation function in order to make the model classify effectively by introducing non-linearity. The fully connected layer was fed into softmax function, which is the final layer in this model.

2.1.2 ResNet152V2

ResNet is referred to as Residual Neural Network, which is useful for image classification tasks. It supports various architectural configurations that helps in attaining satisfactory proportion between work speed and quality. It introduces a structure called residual learning unit, that alleviates the degradation of deep neural networks. The structure of the unit looks like feedforward network, producing best classification accuracy without increasing the model's complexity. Among all the variants of ResNet family, ResNet152V2 is said to have best accuracy.

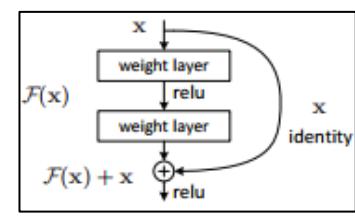


Figure 3: Logical scheme - ResNet

2.1.3 InceptionV3

One of the popular networks called GoogLeNet, is extended further that resulted in InceptionV3. By making use of transfer learning, it achieved greater performance in several applications. InceptionV3 followed GoogLeNet and thereby proposed a new inception model by concatenating multiple different sized convolution filters forming a new filter. This approach of creating a new filter reduces the count of parameters to be involved needs to be trained and thereby decreases the computational complexity.

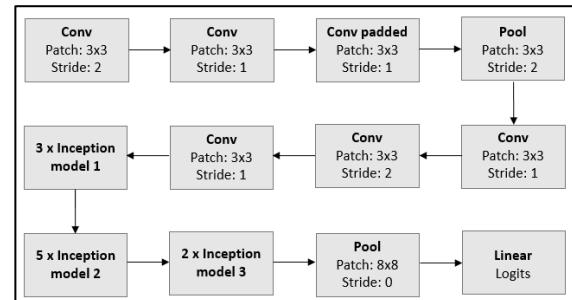
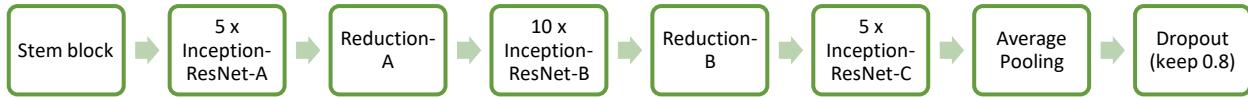


Figure 4: Architecture of Inception-V3

2.1.4 InceptionResNetV2

The combination of Inception structure and Residual connection resulted in the model called InceptionResNetV2. Various multiple sized convolution filters are fused together through residual connections to form an Inception-ResNet block. By using these types of residual connections, degradation problems occurring due to deep structures get avoided, thereby reducing the training time.



2.1.5 DenseNet201

In Dense Convolutional Network (DenseNet), each layer obtains necessary information from all the preceding layers. This information gets passed on to its own feature-maps to all the subsequent layers. Through the process of concatenation, each layer in this model receives “collective knowledge”. Hence, it possesses strong gradient flow, parameter and computational efficiency, diverse features and thereby maintains low complexity features. By utilizing all complexity features, it tends to result in smoother decision boundaries.

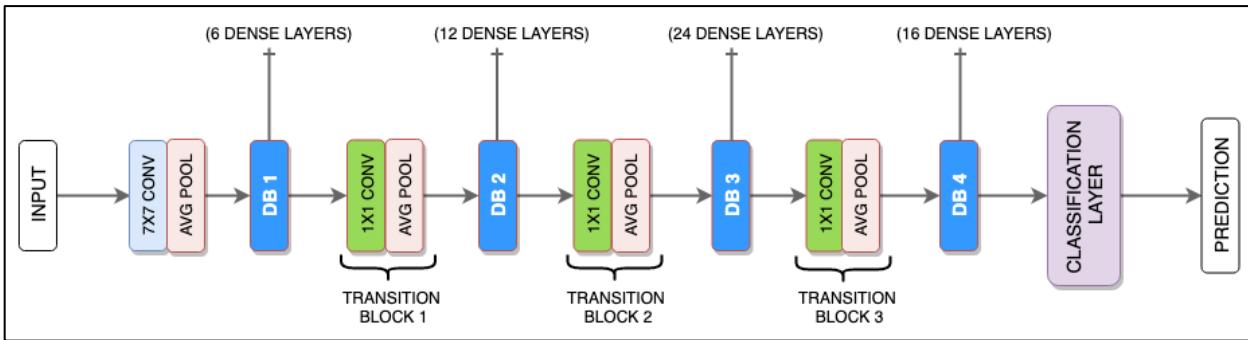
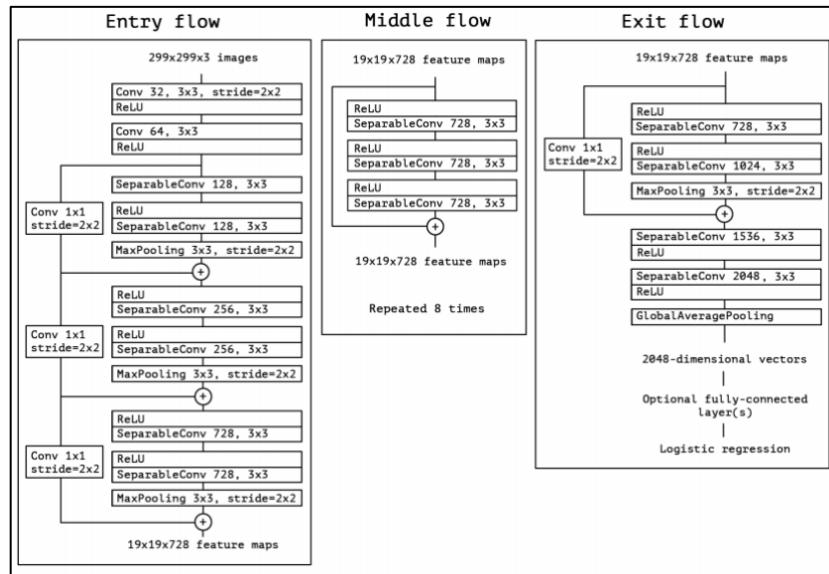


Figure 5: Architecture of DenseNet

2.1.6 Xception

Xception, which is an extension of Inception, replaces the standard inception modules with deepwise separable convolutions. It is one of the convolutional neural networks that contains 71 deep layers.



2.1.7 MobileNetV2

Since, MobileNetV2 uses inverted residual blocks with bottlenecking features, it still differs from with MobileNet with respect to the number of parameters. Having lower parameters count, MobileNetV2 supports any input size which is greater than 32 x 32. It provides much better accuracy, based on the size of the input image. Larger the size of the image, greater the performance.

Two types of blocks exist in MobileNetV2. The major difference between these two blocks is the value of stride. For each block, there are three layers namely convolution layer with ReLU, depth wise convolution layer and convolution layer without non-linearity. Hence, this model is more effective in extracting features on object detection and during the process of semantic segmentation. It is a type of convolutional neural network that outperform on mobile devices.

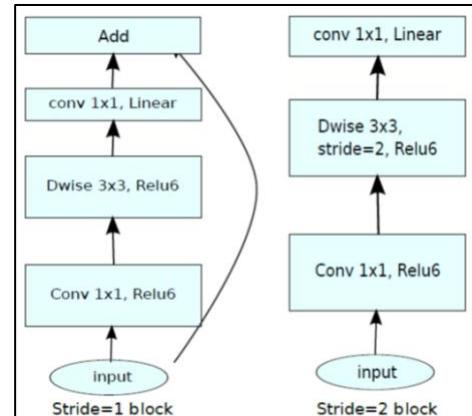


Figure 7: MobileNetV2 Architecture

2.1.8 EfficientNetB5

Efficient Net is said to be the next state-of-the-art network, which has a smaller number of parameters. It has variants ranging from B0 to B7. The resolution of EfficientNetB5 is 456 and training the model is relatively faster. All the existing variants of EfficientNet models are being scaled from Efficient Net-B0 but uses different compound co-efficient. This consistently reduces the number of parameters and FLOPS through an order of magnitude.

2.2 Pixel Based Classification

Each individual pixel is utilized in the classification task in order to a define land cover class, as there exists a difference among each pixel of the imagery. For example, each individual pixel, or group of pixels, will have different brightness values, texture etc., in different bands, which is the major feature that allows the model in predicting the land cover class it originally belongs to. These features and their related statistics derived from each individual pixel allows the trained model to differentiate them.

2.2.1 UNet

UNet, which is one of the popular fully convolutional networks (FCN), comprises of contraction and expansion paths. Contraction path helps in extracting more advanced features and reduces the size of feature maps, whereas expansion path covers the size of segmentation map. UNet is basically a semantic segmentation approach, which is the process of associating each and every pixel of an image, with a corresponding class label.

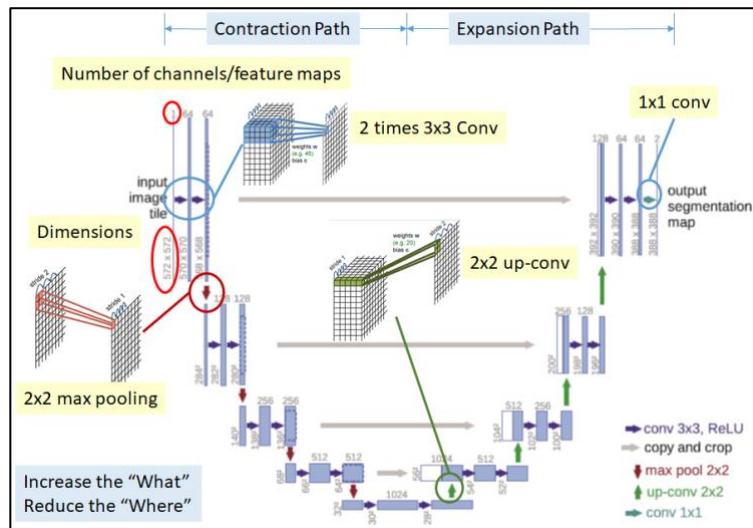


Figure 8: UNet Architecture

3 Study Area and Dataset



Figure 9: JILONA region

The geographic region of interest for this project work is defined as Jupiter Inlet Lighthouse Outstanding Natural Area (JILONA). It is located in the northern Palm Beach County along the Atlantic coast of South Florida.

JILONA is one of the parts of Bureau of Land Management's 27-million-acre National Conservation Lands and it is the only complete unit in east of Mississippi River. It is adjoined by the Loxahatchee River and the Indian River Lagoon and is only half mile away from the Atlantic Ocean. It is situated 14 miles north of West Palm Beach and it is approximately a 2-hour drive from the north of Miami via Interstate 95. The total land cover area of JILONA region is about 120 acres of open space.

The data for this “Deep CNN for LULC Classification using UAS Imagery” project has been collected from very high- resolution drone imagery. The Unmanned Aircraft System (UAS) data used in this project was collected using drone with Micasense Multispectral Sensor Package in two different years. For Scene classification, the three band RGB data collected during 2018 was used, whereas for Pixel classification, five bands Blue, Green, Red, Red edge, Near infrared data collected during September 2020 was being utilized.

Table 1 represents the specification of the sensor used in the aircraft for collecting data in the JILONA region for two separate regions namely JILONA North and JILONA South.

Table 1: RedEdge-MX Camera Specifications (Source: Micasense)

Weight	231.9 g (8.18 oz.) - (Includes DLS 2 and cables)
Dimensions	8.7 cm x 5.9 cm x 4.54 cm (3.4 in x 2.3 in x 1.8 in)
Spectral Bands	Blue, green, red, red edge, near-infrared (NIR) (global shutter, narrowband)
Wavelength (nm)	Blue (475 nm center, 32 nm bandwidth), green (560 nm center, 27 nm bandwidth), red (668 nm center, 16 nm bandwidth), red edge (717 nm center, 12 nm bandwidth), near-IR (842 nm center, 57 nm bandwidth)
RGB Color Output	Global shutter, aligned with all bands
Ground Sample Distance (GSD)	8 cm per pixel (per band) at 120 m (~400 ft) AGL
Capture Rate	One capture per second (all bands), 12-bit RAW
Interfaces	Serial, 10/100/1000 ethernet, removable Wi-Fi, external trigger, GPS, SDHC
Field of View	47.2° HFOV
Triggering Options	Timer mode, overlap mode, external trigger mode (PWM, GPIO, serial, and Ethernet options), manual capture mode



Figure 10: Micasense RedEdge Camera

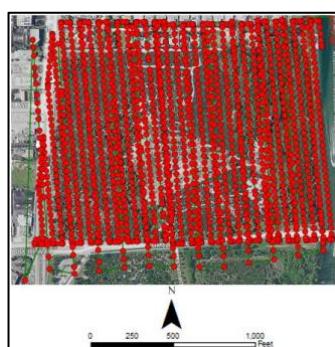


Figure 11: JILONA North

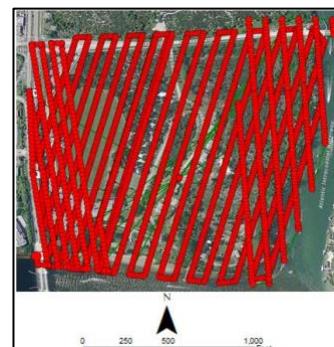


Figure 12: JILONA South

4 Scene Classification

4.1 Training

The drone data collected in the year 2018 was used for scene classification method, which consists of only three band information namely Red, Green, and Blue. Both the North and South imagery was being merged in order to perform scene-based classification. Initially, training samples were collected from the merged imagery of North and South region of JILONA. A total of 9 classes were chosen namely, cabbage palm, grassland, ground, mangrove, parking lot, road, sand pine, scrub oak and sea grape. Figure 13 shows the verified training samples with respect to the ground truth.

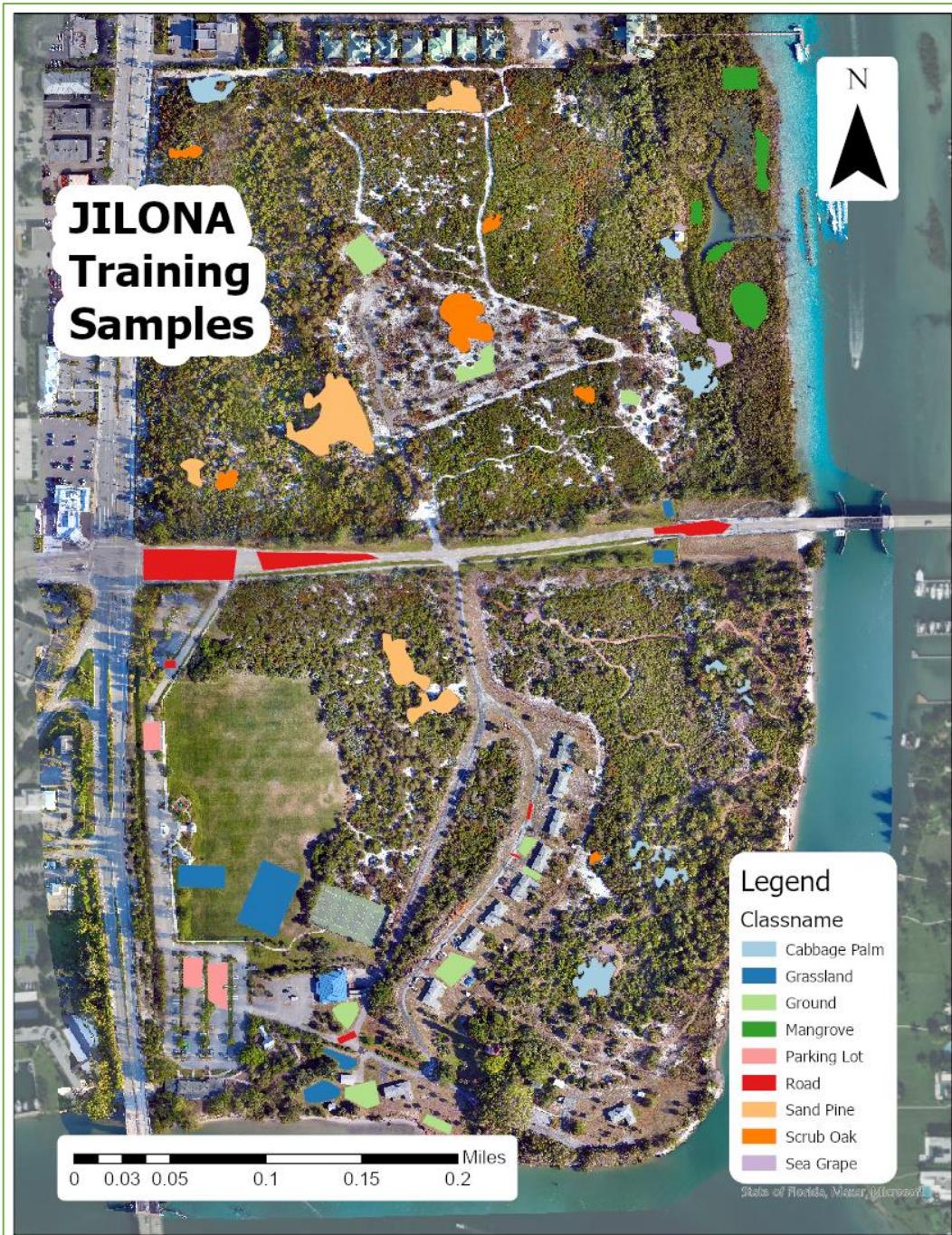


Figure 13: Training samples map - Scene classification

All the collected training samples were then subjected to certain augmentation techniques such as, rescaling, horizontal flipping, vertical flipping, rotation range, brightness range, shear range, and zoom range in order to enlarge the dataset. Image augmentation is a technique in which existing data gets altered, so that more amount of data samples gets generated for model training process. Some of the results of augmentation process is shown in Figure 14. Finally, the whole dataset was split accordingly, so that 80% was allocated to training, 10% for validation and the remaining 10% for testing purposes. The model was trained for about 50 epochs.



Figure 14: Augmented scenes

4.2 Validation and Testing

Once the model is trained using the extracted dataset, it needs to be validated and tested. For both training and testing purposes, **256*256** image patch size was used. Initially, the model was trained using **CPU**, which nearly took many days and hence GPU was set up. Figure 15 explains the time taken for training each scene classification model in the **GPU** desktop utilizing the Tesla T4. The average accuracy of all the models is nearly **93%**.

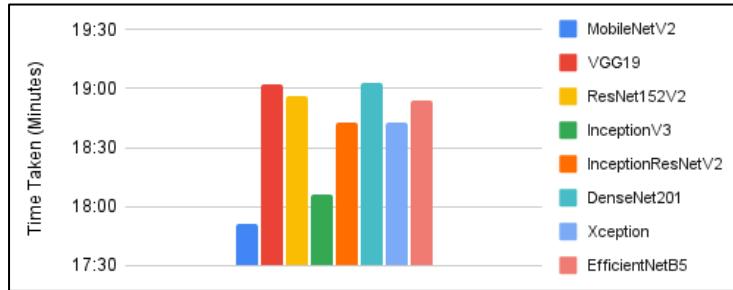


Figure 15: Training time for scene classification method

4.3 Evaluation Metrics

In order to evaluate the developed scene classification models, certain evaluation metrics have been carried which is discussed in detail in this section. This is done to measure the quality of any deep learning model or statistical analysis. The types of evaluation metrics include classification accuracy, confusion matrix, logarithmic loss, Cohen's kappa co-efficient etc.,

The ratio of number of correct predictions to the total number of input data samples is called **classification accuracy**, whereas the false classifications are often referred to as **log loss** or **logarithmic loss**. **Confusion matrix** describes the complete performance of the machine learning model. Evaluation metrics helps in determining whether the model is operating correctly and optimally.

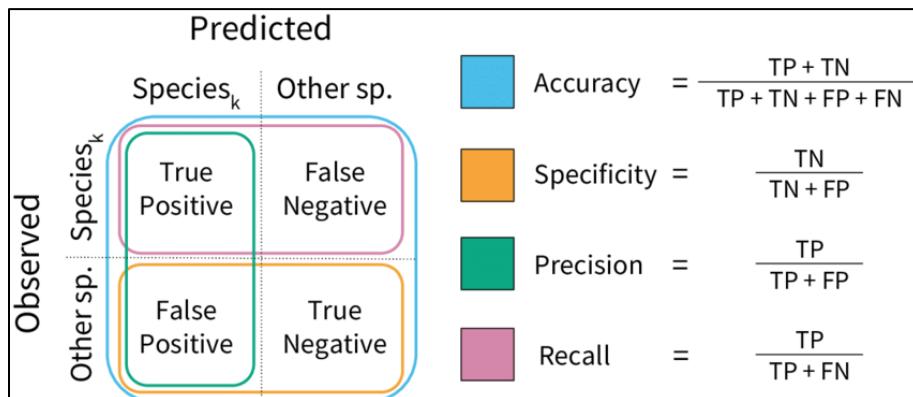


Figure 16: Model performance metrics

$$\kappa = \frac{(p_o - p_e)}{(1 - p_e)}$$

where p_o is the observed proportion of agreement
and p_e is the proportion expected by chance.

Figure 17: Cohen's kappa formula

4.3.1 Accuracy and Loss Charts

The accuracy, validation accuracy, loss, and validation loss for all the explored models can be visualized clearly in Figure 18. Through these charts, one can determine whether the model is learning properly, or it has any issues such as overfitting or diverse probability values, underfitting, cramming values etc., so that model can be insisted to minimize its loss and thereby maximizing the accuracy.

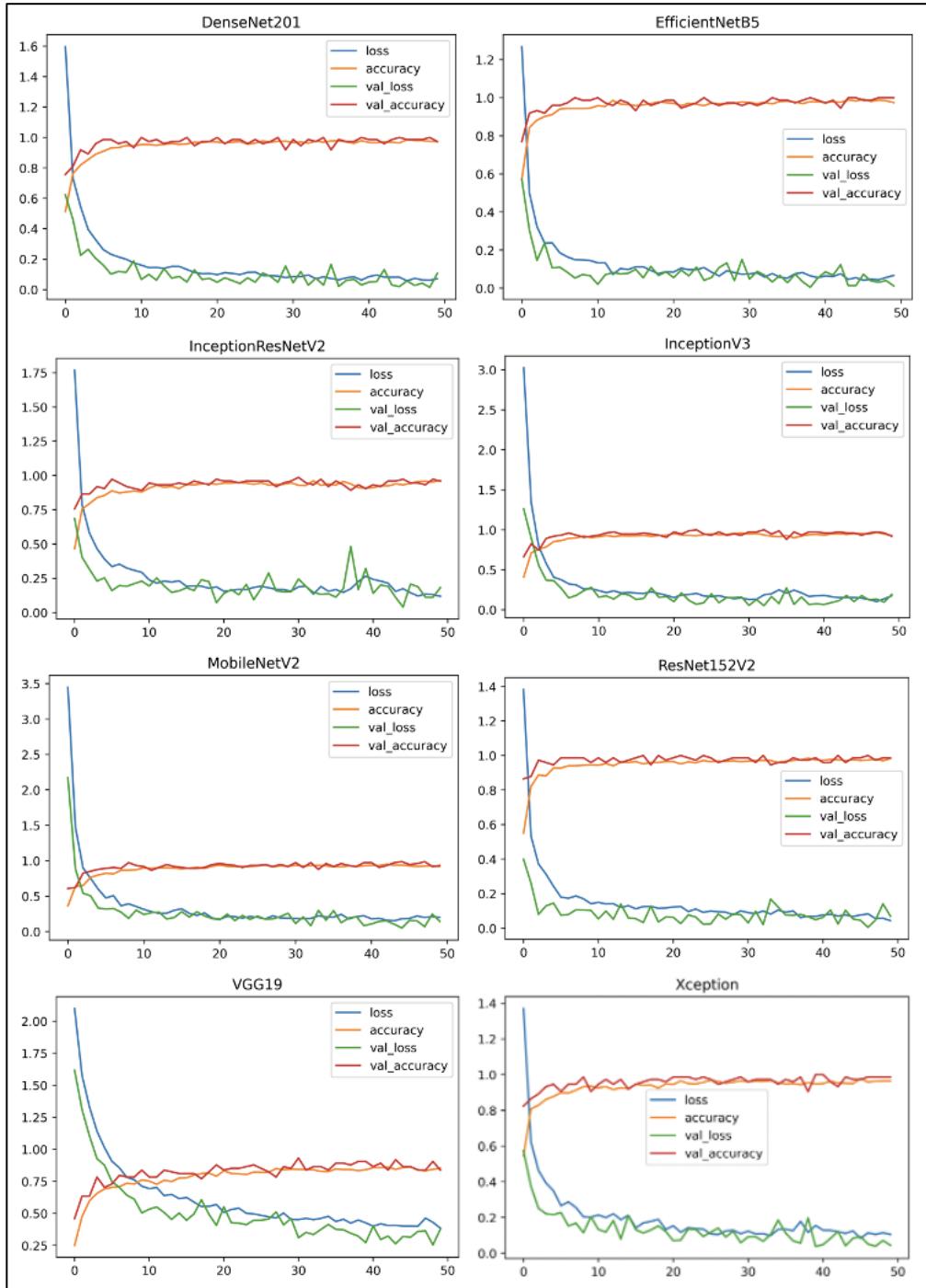


Figure 18: Accuracy and loss charts for scene-based method

4.3.2 Confusion Matrix

EfficientNetB5 and DensNet201 is proven to be the best scene classification model for JILONA region, with an accuracy of nearly 98%. In contrast to that, VGG19 scored the least accuracy percentage of 85.5%. In the same way MobileNetV2, ResNet152V2, InceptionV3, InceptionResNetV2, Xception has accuracy of about 90.6%, 95.1%, 91%, 91.8%, 95.4% respectively.

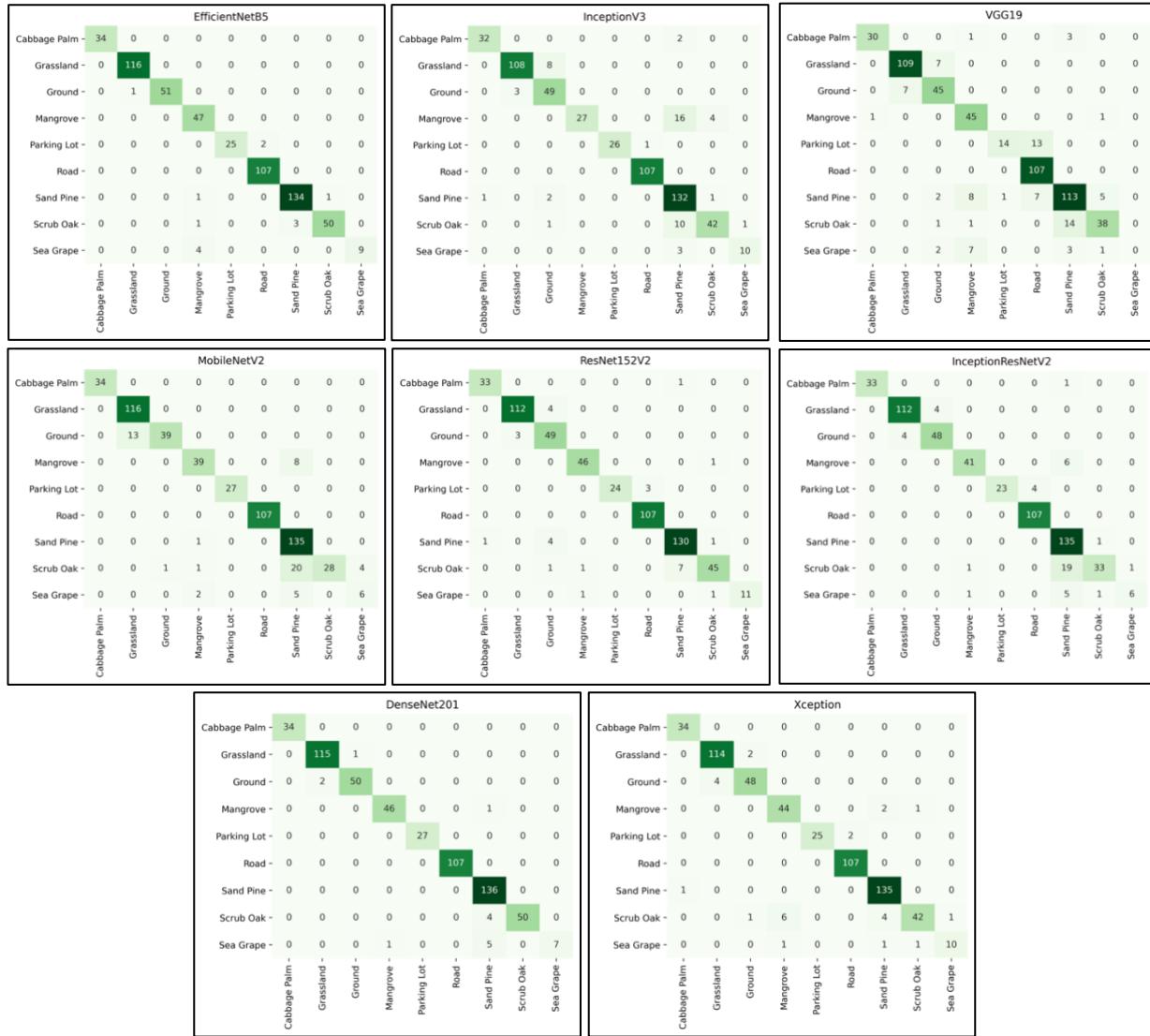


Figure 19: Confusion matrices for all scene-based models

4.3.3 Model Accuracy and Kappa co-efficient

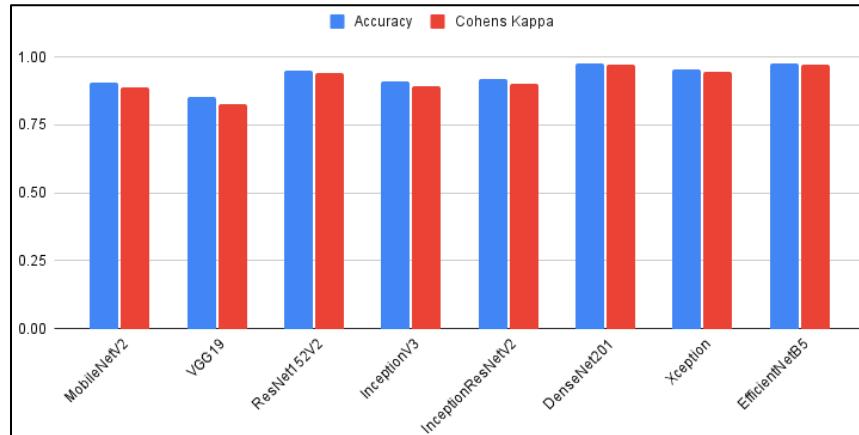


Figure 20: Accuracy, kappa co-efficient for scene classification

5 Pixel Classification

5.1 Training

Nearly 30% of the training data has been collected separately for North and South region of JILONA through ArcGIS Pro from the existing output of Support Vector Machine (SVM) classification since a classification map is needed as an input for the UNet methodology. This is because, the training data was not proper when the mosaicked image of JILONA region is used for training the model. SVM classifier is a supervised learning method, which is being effective in high dimensional spaces and is more memory efficient. The existing datasets are then artificially expanded through a process called “Image augmentation”, to train the deep learning model. Pixel classification is studied using “semantic segmentation” in this research project. Seven classes were considered for South region of JILONA, whereas eight classes for North. The common seven classes are Oak scrub, sand pine, palm, ground, grass, shadow, and water, while mangrove is the additional class in North part of JILONA. The whole idea of this approach is to create a better classification result for the whole region of JILONA than the existing SVM results. With the help of the generated results the UNet model has been developed and trained for various epochs using different approaches of band combinations.



Figure 21: 7 classes
- South JILONA



Figure 22: 8 classes
- North JILONA

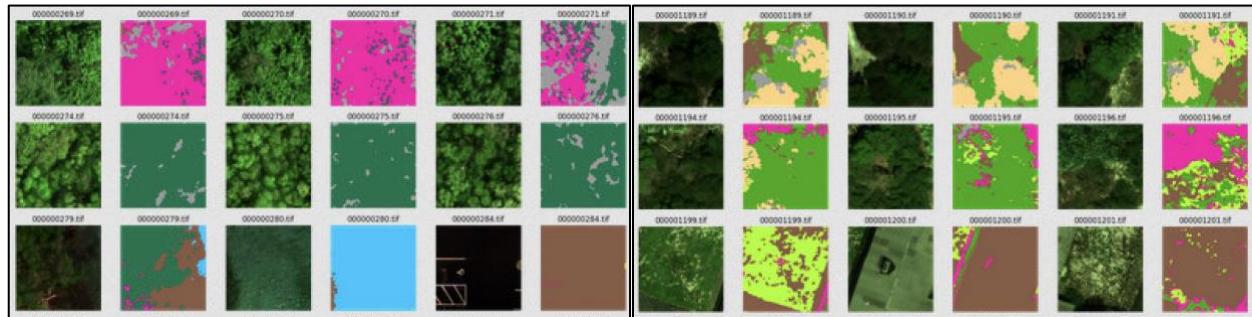


Figure 23: North (left) and South (right) training data

The three chosen combination of band are (1) Red, Green and Blue; (2) Green, Red, Near infrared; (3) Blue, Green, Red, Red edge, Near infrared. Several iterations of model were trained in order to determine the best model for semantic segmentation in the JILONA region. Different set of epochs and hyperparameter tuning were investigated during this study. During the initial phase, model training was performed using CPU, for which it took a lot of days. At latter part of the project, GPU Tesla T4 was utilized to train the model, that takes less than 27 hours for a model to be trained. Meanwhile, Quality Assurance Quality Check (QA-QC) was performed manually for both the North and South training dataset, before using it as input for the model. QA-QC helped in removing a lot of misclassified images, which improved the prediction accuracy.

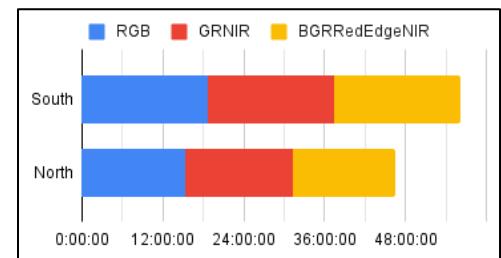


Figure 24: Training time for North
and South region of JILONA

5.2 Experiments

Experiments were carried out to determine the best performing UNet model. Two types of UNet models namely, CustomUNet and SatelliteUNet were examined during this study with three different band combinations. Certain other hyper parameters such as number of layers, learning rate, batch size, optimizer was also detailedly investigated in this research work. In addition to that, it is also proven that the models outclass on using augmented images for training North and South region of JILONA separately. Several epochs of training were carried out to determine the improvement of model.

All the assessments that were made during this study is explained in Figure 26. The best performing model, out of all demonstrations is also highlighted in Figure 26.

Parameters	Investigated values	Best Model
Models	CustomUNet, SatelliteUNET	CustomUNet
Band Combinations	RGB, GRNIR, BGRRedEdgeNIR	BGRRedEdgeNIR
Number of Layers	4, 5, 6	5
Learning Rate	0.01, 0.001, 0.0001	0.01
Batch Size	16, 32, 64	64
Optimizer	Adam, SGD	Adam
Epochs	100, 500, 1000, 1500, 2000, 2500	2500
Augmentation	Yes, No	Yes
Image	Mosaicked, North-South separately	North-South separately

Figure 25: Investigated values for UNet model

5.3 Validation and Testing

Once the training is completed for all the developed models, testing was performed for randomly chosen points, which has ground verified class. The testing samples of North and South region of JILONA can be visualized in the maps shown in Figures 27 and 28. Individual accuracy of each class in every model is shown in figures 29, 30, 31, 32, 33 and 34.

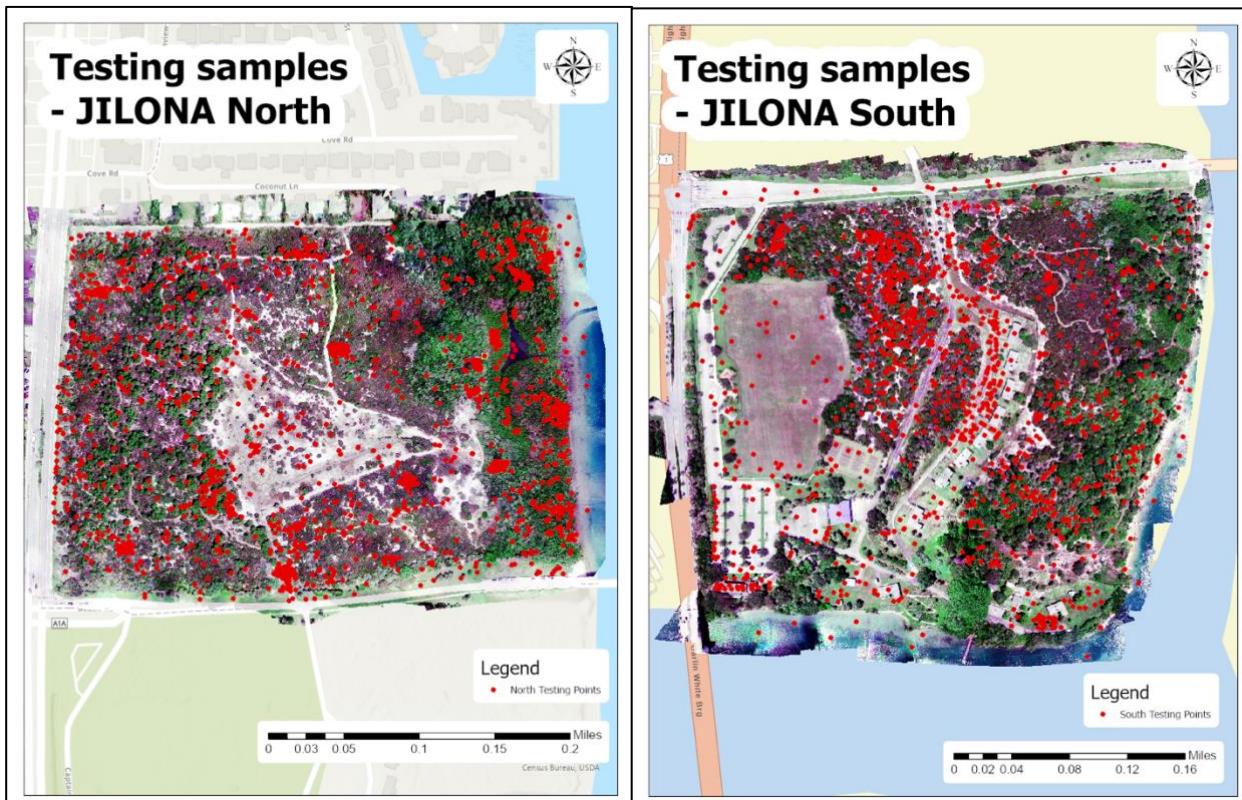


Figure 26: North JILONA testing samples

Figure 27: South JILONA testing samples

	Total	Pass	Pass%
Ground	224	203	90.62
SandPine	197	180	91.37
OakScrub	236	220	93.22
Grass	218	210	96.33
Palm	266	243	91.35
Shadow	69	44	63.77
Water	15	15	100.00

Figure 28: South All 5 Bands

	Total	Pass	Pass%
Ground	224	197	87.95
SandPine	197	178	90.36
OakScrub	236	200	84.75
Grass	218	210	96.33
Palm	266	239	89.85
Shadow	69	39	56.52
Water	15	15	100.00

Figure 29: South - RGB

	Total	Pass	Pass%
Ground	224	204	91.07
SandPine	197	176	89.34
OakScrub	236	212	89.83
Grass	218	217	99.54
Palm	266	236	88.72
Shadow	69	18	26.09
Water	15	15	100.00

Figure 30: South - GRNIR

	Total	Pass	Pass%
SandPine	279	272	97.49
OakScrub	288	201	69.79
Palm	251	225	89.64
Ground	316	291	92.09
Water	31	28	90.32
Grass	82	29	35.37
Mangrove	398	378	94.97
Shadow	134	124	92.54

Figure 31: North - All 5 bands

	Total	Pass	Pass%
SandPine	279	276	98.92
OakScrub	288	237	82.29
Palm	251	209	83.27
Ground	316	264	83.54
Water	31	28	90.32
Grass	82	40	48.78
Mangrove	398	373	93.72
Shadow	134	120	89.55

Figure 32: North - RGB

	Total	Pass	Pass%
SandPine	279	274	98.21
OakScrub	288	200	69.44
Palm	251	226	90.04
Ground	316	279	88.29
Water	31	28	90.32
Grass	82	32	39.02
Mangrove	398	365	91.71
Shadow	134	124	92.54

Figure 33: North - GRNIR

5.4 Evaluation Metrics

Overall accuracy for three major results of this study and the currently existing output of SVM model are being discussed in Table 2. It is evident that, the Custom UNet model outperforms while using all the five-band information with 87.02% of accuracy for North imagery and 91% of accuracy for South imagery of JILONA region. Using the Equation 1, the overall accuracy was calculated based on the individual accuracies explained in section 5.3.

$$\text{Overall Accuracy} = \frac{\text{Number of correctly classified pixels}}{\text{Total number of pixels}} \quad (1)$$

Table 2: Accuracy comparison of existing and proposed model

METHODOLOGY	NORTH ACCURACY	SOUTH ACCURACY
Custom UNet - RGB	86.96%	88%
Custom UNet - GRNIR	85.89%	88%
Custom UNet - BGRRedEdgeNIR	87.02%	91%
SVM - Without DEM	74%	84%
SVM - With DEM	89%	91%

5.4.1 Accuracy and Loss Charts

Accuracy and loss charts for the Custom UNet models for all three band combinations is shown in figures 35, 36 and 37. For all the five-band combination, 2000 epochs of training are chosen to be the best, whereas for three band combinations, training for datasets is proven good for 2500 epochs.

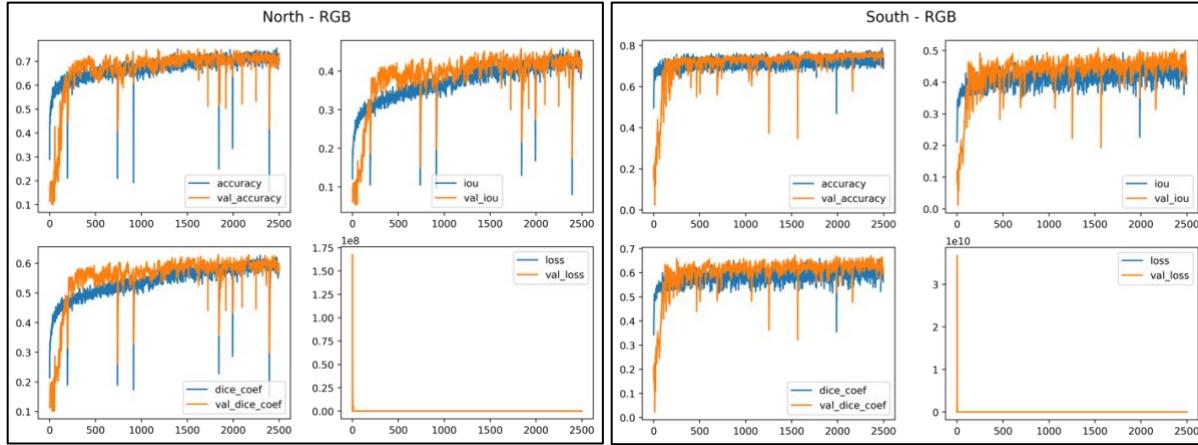


Figure 34: Accuracy and loss charts of North, South - RGB band

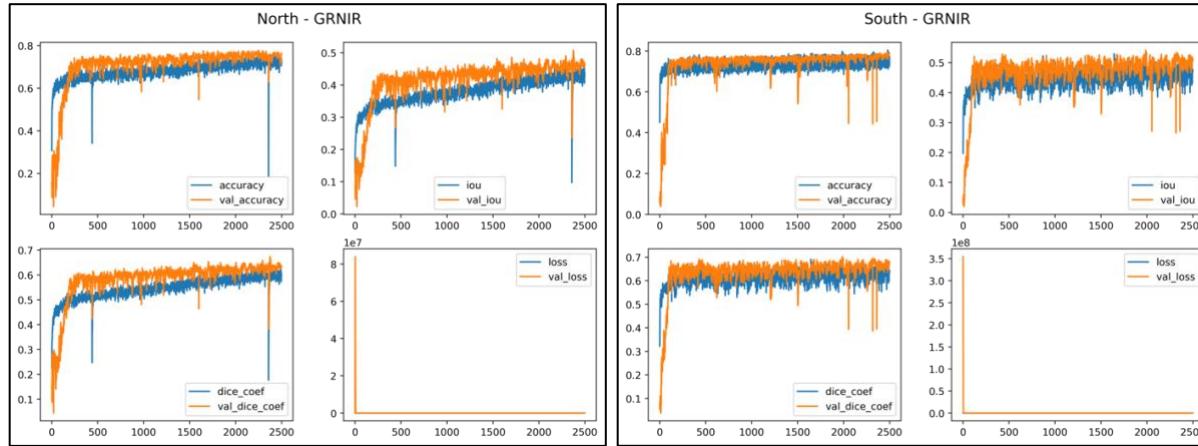


Figure 35: Accuracy and loss charts of North, South - GRNIR band

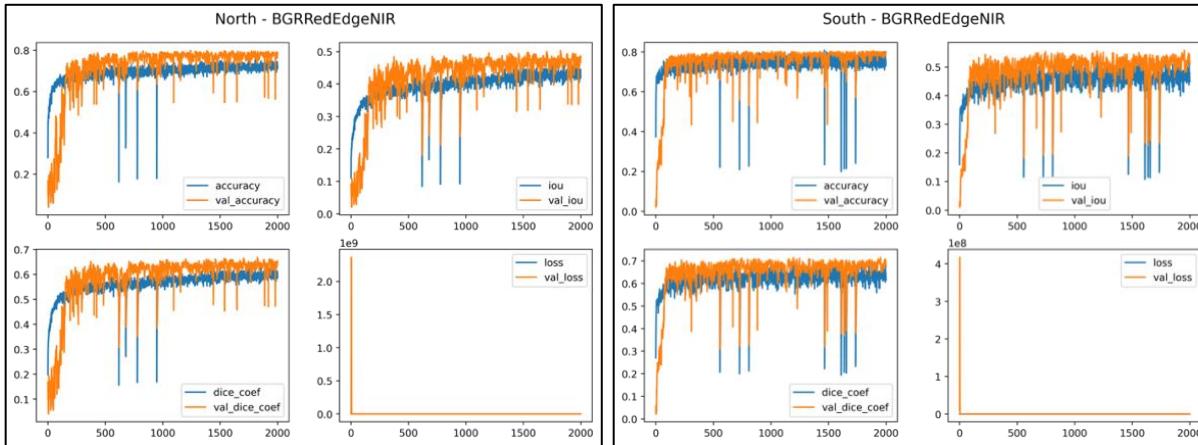


Figure 36: Accuracy and loss charts of North, South - BGRRedEdgeNIR band

5.4.2 Classification map



Figure 37: South - RGB

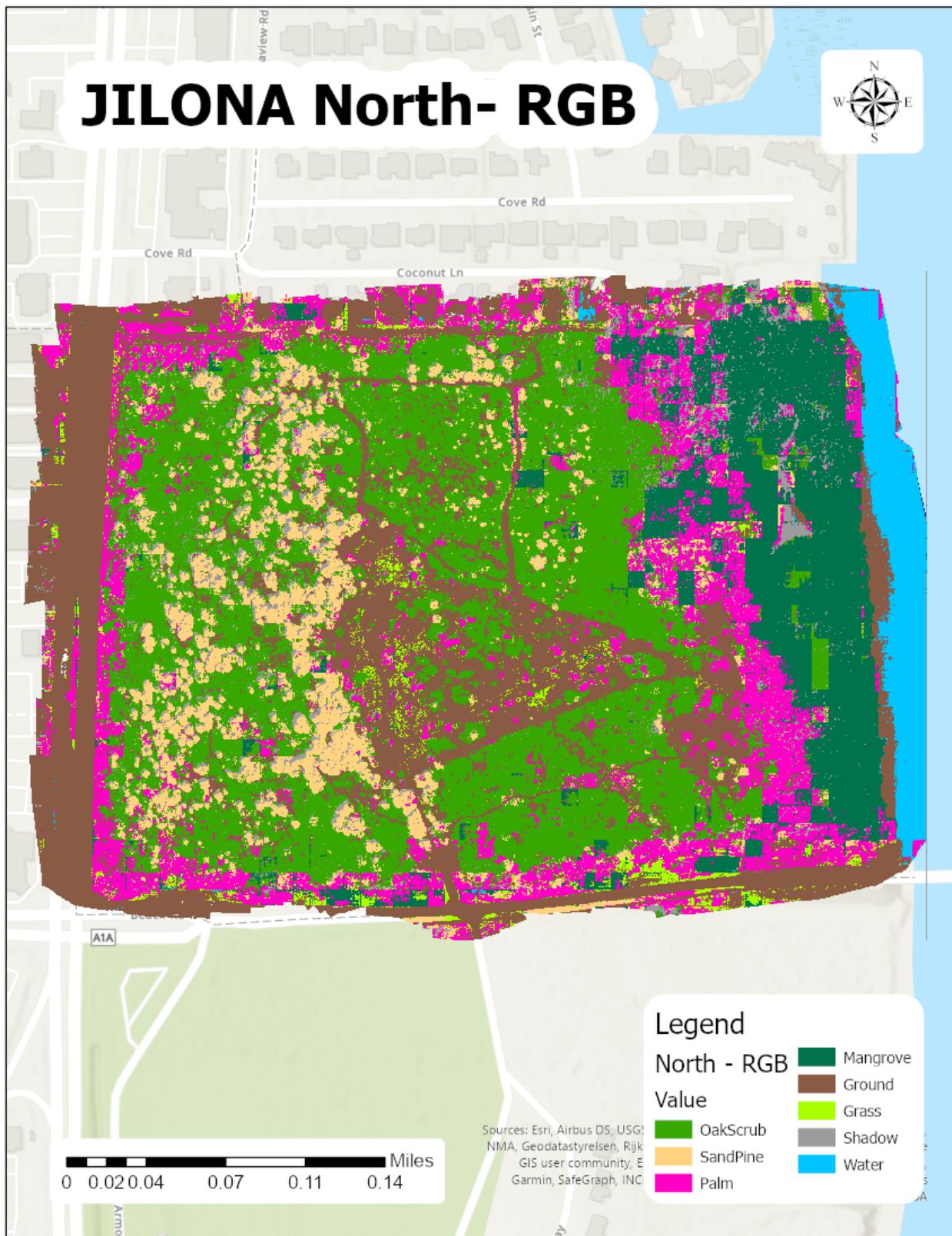


Figure 38: North - RGB



Figure 39: South - GRNIR

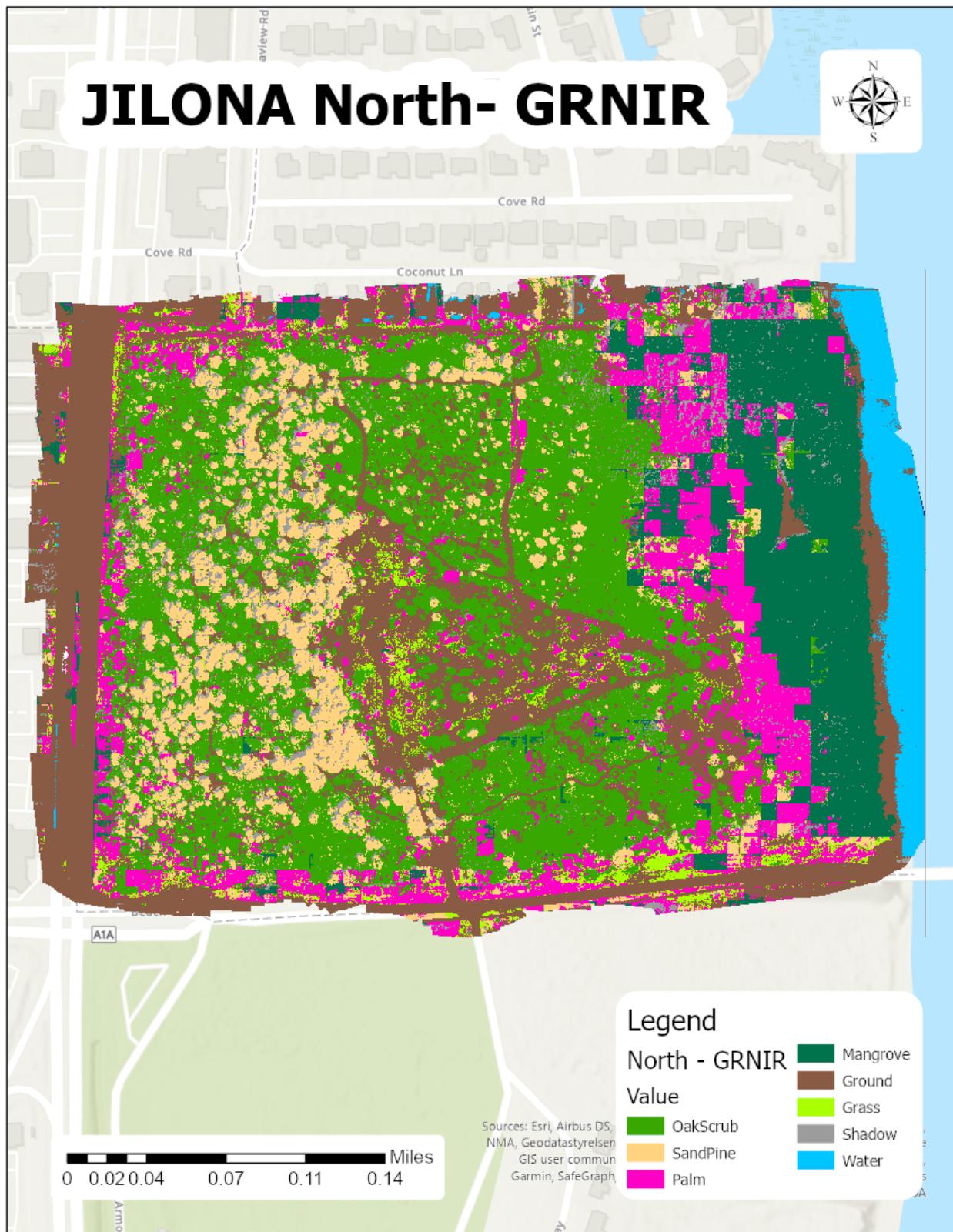


Figure 40: North - GRNIR



Figure 41: South - All 5 bands



Figure 42: North - All 5 bands

6 In-depth comparisons of UNet and SVM results

6.1 South region of JILONA

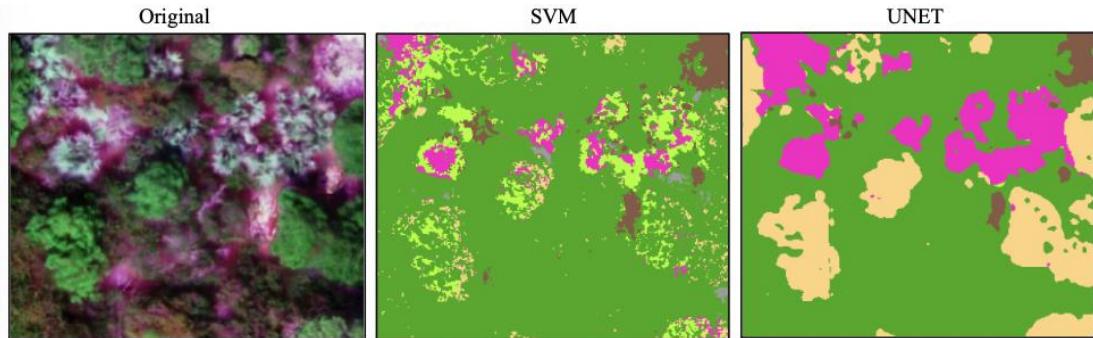


Figure 43: Correctly classified Cabbage palm (pink) and Sand pine (yellow)

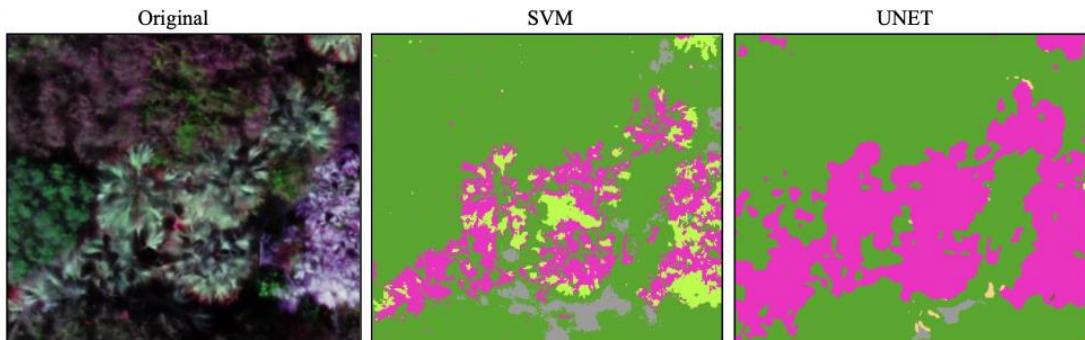


Figure 44: Correctly classified Cabbage palm (pink) and Grass (light green)

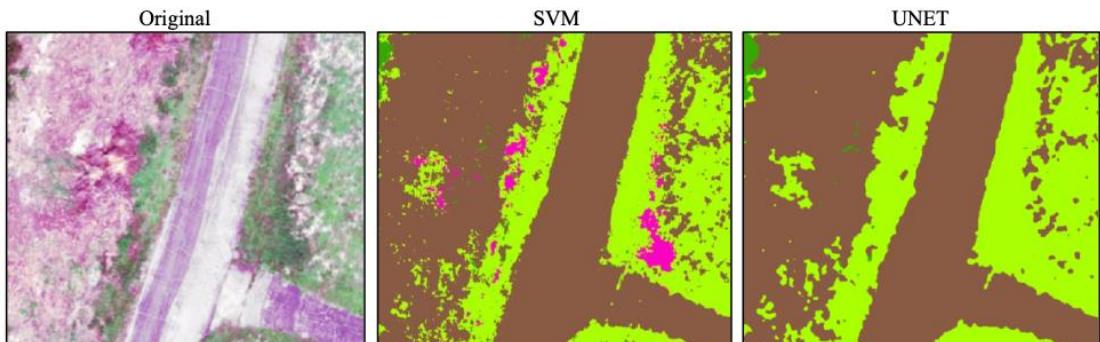


Figure 45: Correctly classified Ground (brown) and Grass (light green)

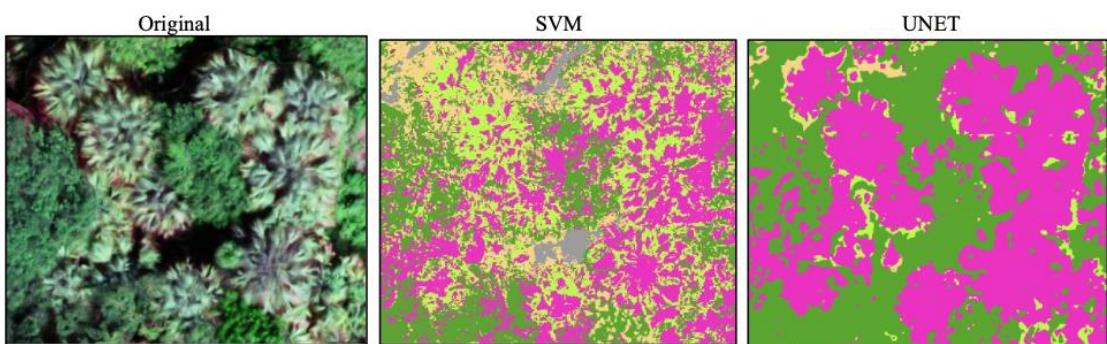


Figure 46: Correctly classified Cabbage palm (pink)

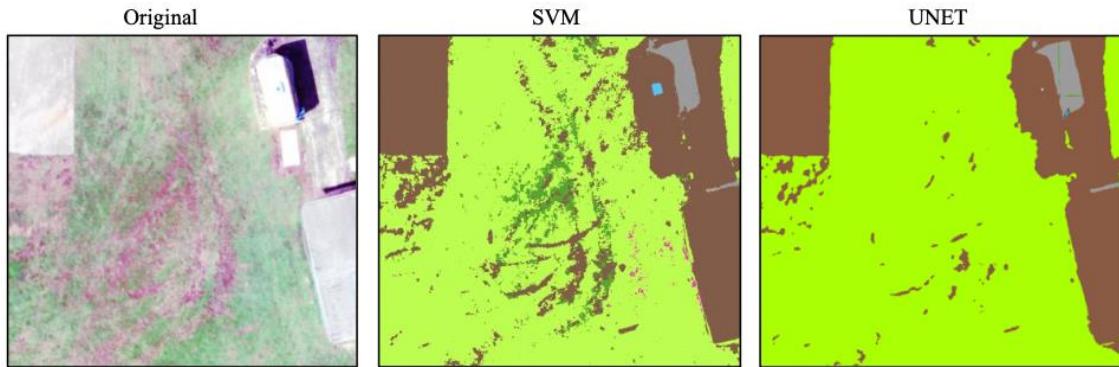


Figure 47: Correctly classified Grass (light green)

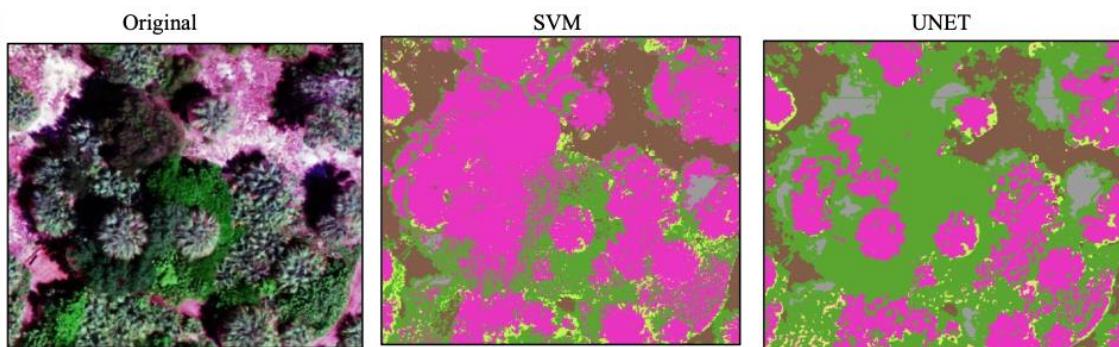


Figure 48: Correctly classified Oak scrub (dark green)

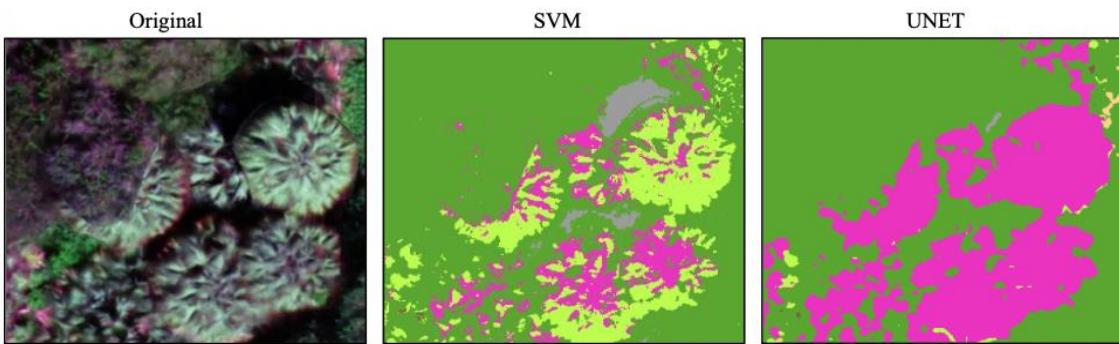


Figure 49: Correctly classified Cabbage palm (pink)

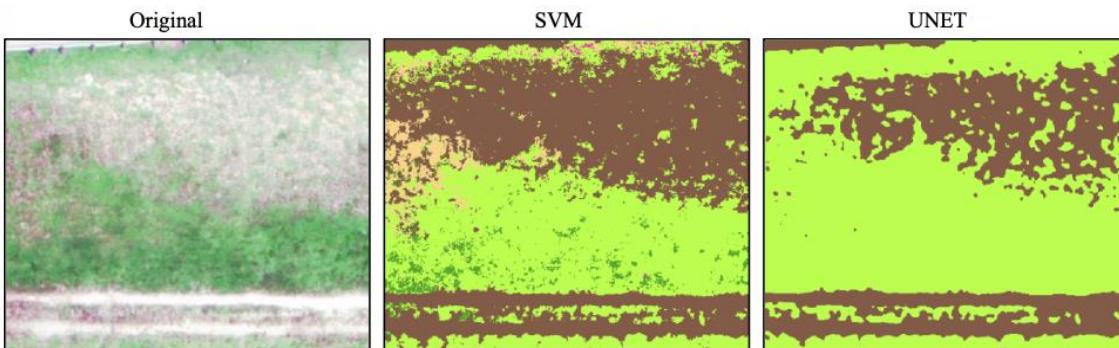


Figure 50: Correctly classified Grass (light green) and Ground (brown)

6.2 North region of JILONA

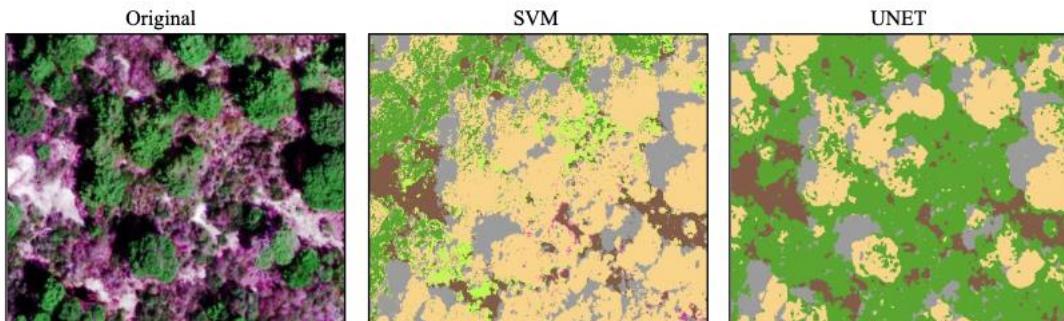


Figure 51: Correctly classified Oak scrub (dark green) and Sand pine (yellow)

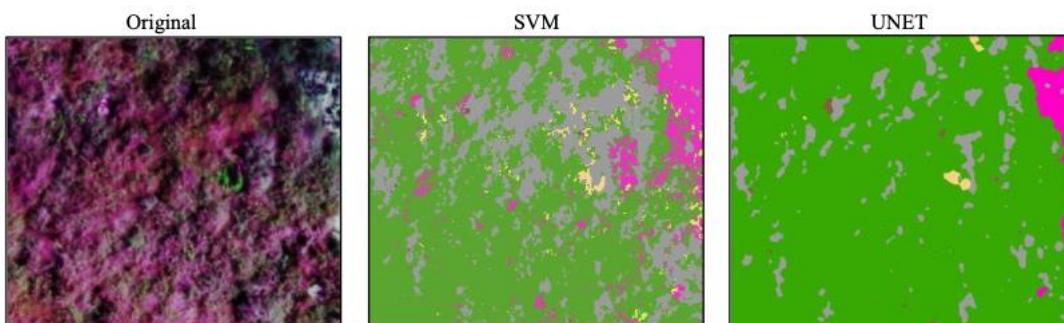


Figure 52: Correctly classified Oak scrub (dark green)

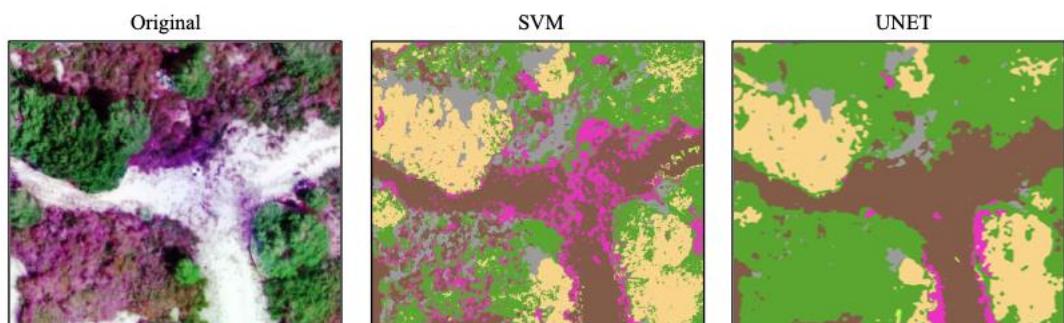


Figure 53: Correctly classified Ground (brown)

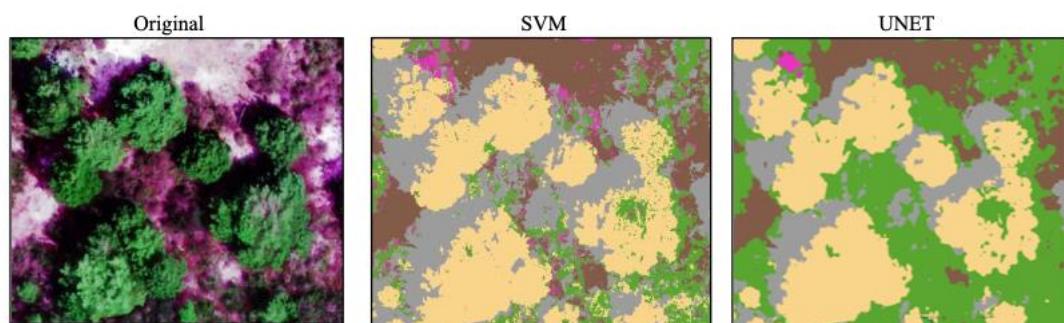


Figure 54: Correctly classified Oak scrub (dark green)

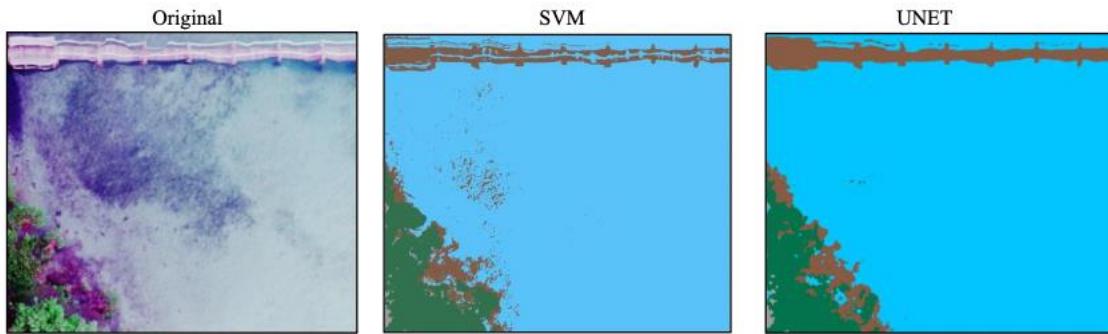


Figure 55: Correctly classified Ground (brown) and water (blue)

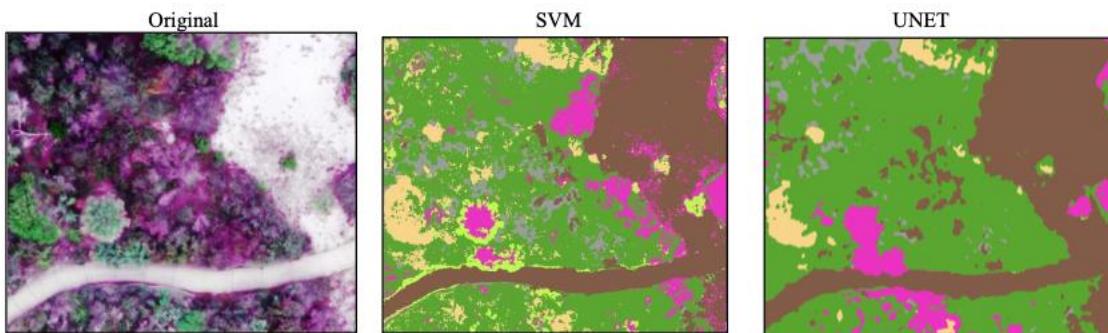


Figure 56: Correctly classified Ground (brown) and Cabbage palm (pink)

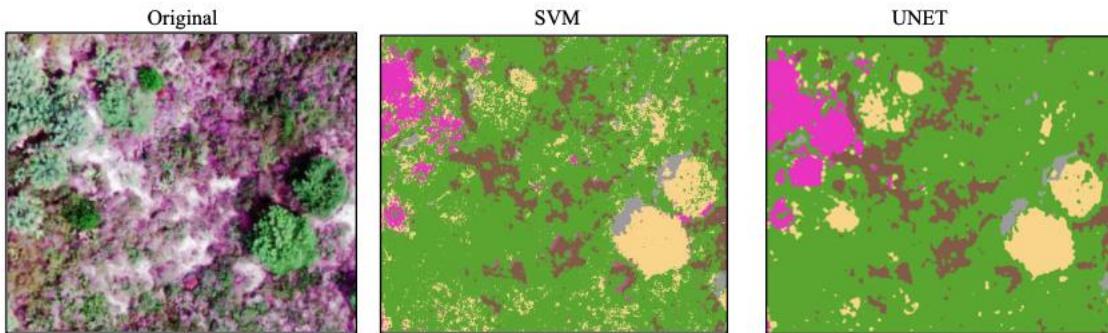


Figure 57: Correctly classified Cabbage palm (pink) and Oak scrub (dark green)

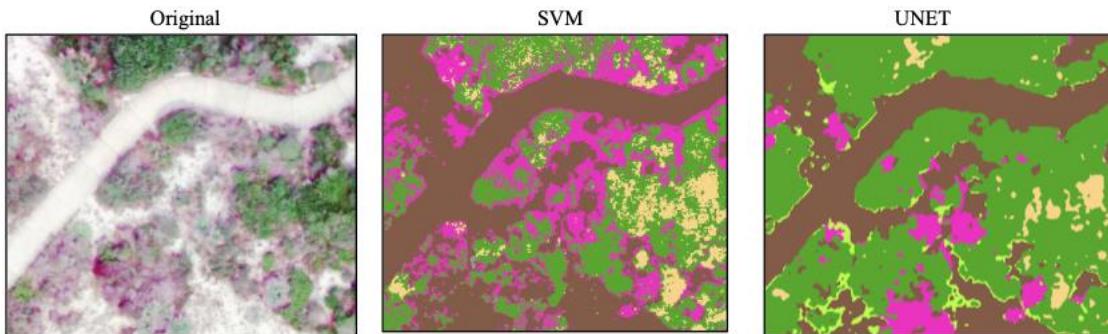


Figure 58: Correctly classified Ground (brown) and Cabbage palm (pink)

7 Detected problems in UNet

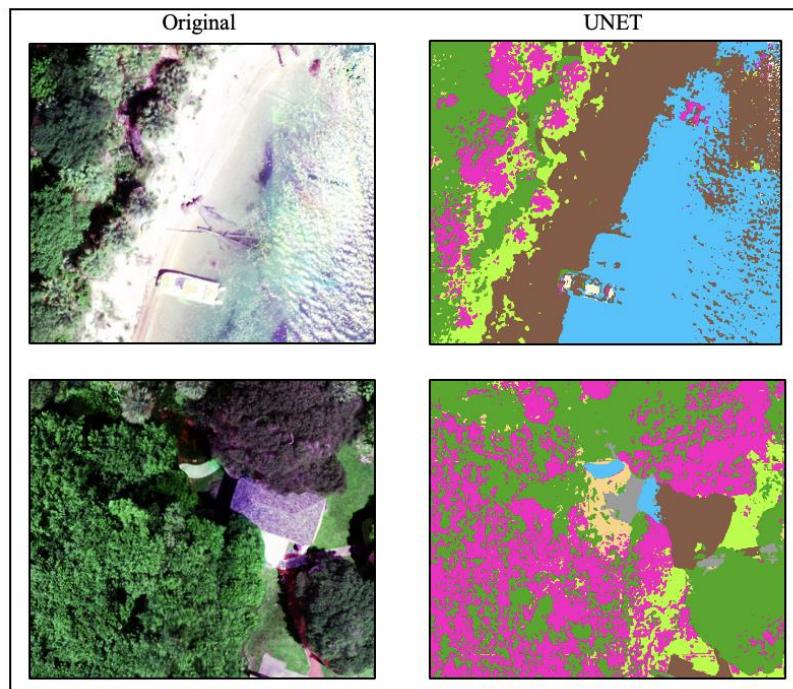


Figure 59: Misclassified Water (blue), Ground (brown), Ground (brown) and Cabbage palm (pink) - South region of JILONA

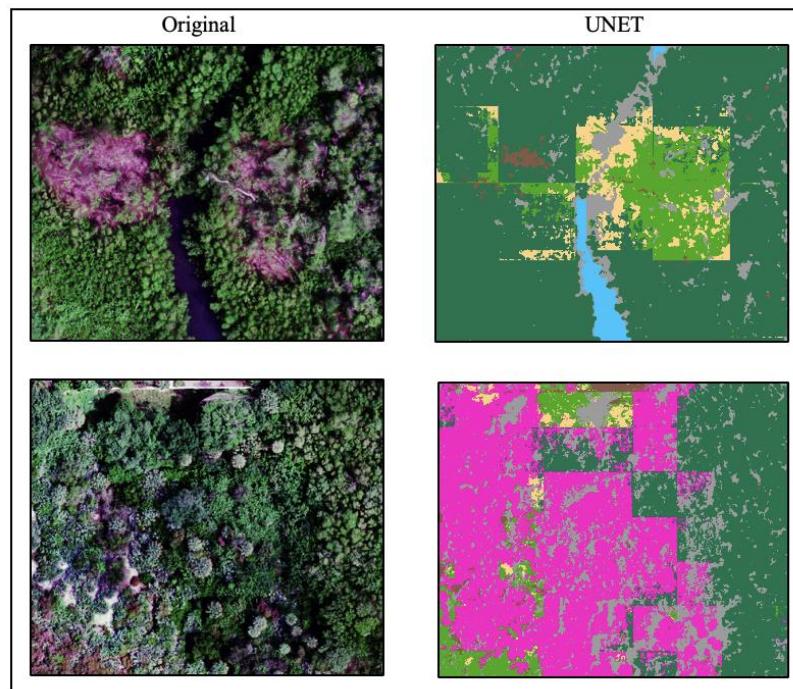


Figure 60: Misclassified Sand pine (yellow), Cabbage palm (pink) and Mangrove (deep dark green) - North region of JILONA

8 Conclusion

Scene Classification and Semantic Segmentation were performed for JILONA region. The performance of 8 scene-based models namely VGG19, ResNet152V2, InceptionV3, EfficientNetB5, Xception, InceptionResNetV2, MobileNetV2, DenseNet201 were compared in this study. In addition, various combinations of hyperparameters were tuned in order to capture the finest working model for pixel-based classification. Finally, Land use - Land cover classification map was generated from the determined best UNet segmentation model. This model was developed using the UAS imagery, which is cost effective as it utilizes a low-cost sensor.

9 Results and Discussion

The finding of this research is that among all the evaluated models, **EfficientNetB5** outperformed in scene classification with nearly 98% of prediction accuracy. Similarly, **CustomUNet** with all **5-band combination** trained using Adam optimizer, 64 as batch size, 5 layers and 0.01 learning rate, outclassed all other investigated models for pixel-based classification or semantic segmentation. One important thing to be noted is that accuracy percentage of UNet model, got elevated after performing the Quality Inspection for training images.

10 Future Scope

The current study is restricted to a specific study area, which can be extended in the future. Also, the comparison of the classification methods can be carried out for various other band combinations of the same (or) different sensors. Further, an in-depth analysis can be performed in order to improve the prediction accuracy of the UNet model. Similarly, hyperparameters can be tuned in such a way, so that the scene classification model improves. Moreover, there are some patch issues in the classification map, which can be seen in Figure 61. In future, the model needs to be trained, so that this issue gets resolved.

11 Acknowledgements

The author would like to thank Bureau of Land Management for providing necessary support and wishes to express sincere gratitude to guidance and support of supervisor Dr. Sudhagar Nagarajan, and a special thanks to the Department of Civil, Environmental and Geomatics Engineering at Florida Atlantic University for allowing to carry out the research. In addition, the author is much grateful to all colleagues for academic assistance.

12 Author Contributions

Conceptualization, MR and Dr. SN.; Analysis and interpretation of result, MR; Investigation, Dr. SN; Methodology, MR; Supervision, Dr. SN; Validation, MR and Dr. SN; Report writing, MR.

13 Funding

Though this research is an outcome of Directed Independent Study, it used the data collected through the funding provided by Bureau of Land Management - JILONA.

14 Conflicts of Interest

The authors declare no conflict of interest with any other researchers or entities.

15 References

- 1) Adão, Telmo, et al. "Hyperspectral Imaging: A Review on UAV-Based Sensors, Data Processing and Applications for Agriculture and Forestry." *Remote Sensing*, vol. 9, no. 11, 2017, p. 1110., doi:10.3390/rs9111110.
- 2) Abdishakur: Land use/Land cover classification with Deep Learning. [https://towardsdatascience.com/land-use-land-cover-classification-with-deep-learning-9a5041095ddb 2018](https://towardsdatascience.com/land-use-land-cover-classification-with-deep-learning-9a5041095ddb). (Accessed 24 July 2021).
- 3) Bannari, A., et al. "A Review of Vegetation Indices." *Remote Sensing Reviews*, vol. 13, no. 1-2, 1995, pp. 95–120., doi:10.1080/02757259509532298.
- 4) Bradski, G.: The OpenCV Library. In: Dr. Dobb's Journal of Software Tools (2000).
- 5) Chollet, Francois, Keras Applications. Available online: <https://keras.io/api/applications/> (Accessed on 23 June 2021).
- 6) Christopher Benjamin Anderson. The CCB-ID approach to tree species mapping with airborne imaging spectroscopy. <http://dx.doi.org/10.7717/peerj.5666> (Accessed on 23 July 2021).
- 7) Cohen's kappa co-efficient. Available online: <http://lboringoldman.com/index.php/2012/09/24/take-notic/> (Accessed on 22 July 2021).
- 8) Data Science - Available online: <https://towardsdatascience.com/review-densenet-image-classification-b6631a8ef803> (Accessed on 16 July 2021).
- 9) Deep Learning. Available online: <https://www.deeplearningbook.org/> (Accessed on 5 June 2021).
- 10) Deep Learning with Python by François Chollet (Accessed on 21 June 2021).
- 11) DenseNet Model. Available online: <https://towardsdatascience.com/paper-review-densenet-densely-connected-convolutional-networks-acf9065dfefb> (Accessed on 23 July 2021).
- 12) DenseNet-201 - Available online: <https://www.mathworks.com/help/deeplearning/ref/densenet201.html> (Accessed on 20 July 2021).
- 13) Evaluation Metrics. Available online: <https://deeppai.org/machine-learning-glossary-and-terms/evaluation-metrics> (Accessed on 21 July 2021).
- 14) Foody, Giles M. "Sample Size Determination for Image Classification Accuracy Assessment and Comparison." *International Journal of Remote Sensing*, vol. 30, no. 20, 2009, pp. 5273–5291., doi:10.1080/01431160903130937.
- 15) Gillies, Sean u.a.: Rasterio: geospatial raster I/O for Python programmers. 2013. Available online: <https://github.com/mapbox/rasterio>.
- 16) He, Kaiming ; Zhang, Xiangyu ; Ren, Shaoqing ; Sun, Jian: Deep Residual Learning for Image Recognition. In: arXiv preprint arXiv:1512.03385 (2015). – URL <https://arxiv.org/abs/1512.03385>.
- 17) Hunter, J. D.: Matplotlib: A 2D graphics environment. In: *Computing in Science & Engineering* 9 (2007), Nr. 3, S. 90–95. Available Online: <https://matplotlib.org>.
- 18) Image Augmentation. Available online: <https://www.analyticsvidhya.com/blog/2021/03/image-augmentation-techniques-for-training-deep-learning-models/> (Accessed on 23 July 2021).

-
-
- 19) Josh King, Vayu Kishore, Filippo Ranalli. Scene Classification with Convolutional Neural Networks (CNN). Available online: <http://cs231n.stanford.edu/reports/2017/pdfs/102.pdf> (Accessed on 12 July 2021).
 - 20) Kaiming He, Xiangyu Zhang, et al. Deep Residual Learning for Image Recognition (2015): arXiv:1512.03385v1 (Accessed on 22 July 2021).
 - 21) Long Nguyen, Dongyun Lin, et al. Deep CNNs for microscopic image classification by exploiting transfer learning and feature concatenation (Accessed on 21 July 2021).
 - 22) Lv, Qi, et al. "Urban Land Use and Land Cover Classification Using Remotely Sensed SAR Data through Deep Belief Networks." Journal of Sensors, vol. 2015, 2015, pp. 1–10., doi:10.1155/2015/538063.
 - 23) Masoud Mahdianpari, Yun Zhang, et al. Very Deep Convolutional Neural Networks for Complex Land Cover Mapping Using Multispectral Remote Sensing Imagery: <http://dx.doi.org/10.3390/rs10071119> (Accessed on 22 July 2021).
 - 24) Masoud Mahdianpari, Bahram Saelhi, Mohammad Rezae, Fariba Mohammadimanesh and Yun Zhang. Very Deep Convolutional Neural Networks for Complex Land Cover Mapping Using Multispectral Remote Sensing Imagery, 2018 MDPI, Basel, Switzerland.
 - 25) Milas, Anita Simic, et al. "Different Colours of Shadows: Classification of UAV Images." International Journal of Remote Sensing, vol. 38, no. 8-10, 2017, pp. 3084–3100., doi:10.1080/01431161.2016.1274449.
 - 26) MobileNetV2 model from TensorFlow API of Keras application - Available online: https://www.tensorflow.org/api_docs/python/tf/keras/applications/mobilenet_v2(Accessed on 11 July 2021).
 - 27) Nagarajan, Sudhagar, et al. "UAS Based 3D Shoreline Change Detection of Jupiter Inlet Lighthouse ONA after Hurricane Irma." International Journal of Remote Sensing, vol. 40, no. 24, 2019, pp. 9140–9158., doi:10.1080/01431161.2019.1569792.
 - 28) Neubert, M. & Meinel, G. "Evaluation of segmentation programs for high resolution remote sensing applications." Proceedings of the Joint ISPRS/EARSeL Workshop, 2003.
 - 29) Pandian Kesavan, Dr. Sudhagar Nagarajan, et al. Machine Learning Approach for Vegetation Classification Using UAS Multispectral Imagery: Florida Atlantic University (Accessed on 23 July 2021).
 - 30) Reis, Mariane, et al. "Towards a Reproducible LULC Hierarchical Class Legend for Use in the Southwest of Pará State, Brazil: A Comparison with Remote Sensing Data-Driven Hierarchies." Land, vol. 7, no. 2, 2018, p. 65., doi:10.3390/land7020065.
 - 31) ResNet152: <https://supervise.ly/explore/models/res-net-152-image-net-4228/overview> (Accessed on 22 July 2021).
 - 32) Semantic Segmentation. Available online: <https://paperswithcode.com/task/semantic-segmentation> (Accessed on 25 July 2021).
 - 33) SVM Classifier. Available online: <https://scikit-learn.org/stable/modules/svm.html> (Accessed on 23 July 2021)
 - 34) UNet Architecture. Available online: <https://towardsdatascience.com/review-u-net-biomedical-image-segmentation-d02bf06ca760> (Accessed on 20 July 2021).
 - 35) UNet Segmentation advantages. Available online: <https://jean-charles-nigretto.medium.com/advantages-of-u-net-for-image-segmentation-8ce869d28b4d> (Accessed on 24 July 2021).

-
-
- 36) VGG model, Available online: <https://iq.opengenus.org/vgg19-architecture/> (Accessed on 21 June 2021).
 - 37) Wan, Liang, et al. "Combining UAV-Based Vegetation Indices and Image Classification to Estimate Flower Number in Oilseed Rape." *Remote Sensing*, vol. 10, no. 9, 2018, p. 1484., doi:10.3390/rs10091484.
 - 38) Watts, Adam C., et al. "Unmanned Aircraft Systems in Remote Sensing and Scientific Research: Classification and Considerations of Use." *Remote Sensing*, vol. 4, no. 6, 2012, pp. 1671–1692., doi:10.3390/rs4061671.
 - 39) Xception Model. Available online: <https://www.mathworks.com/help/deeplearning/ref/xception.html> (Accessed on 24 July 2021)
 - 40) Xie, Zhuli, et al. "Classification of Land Cover, Forest, and Tree Species Classes with ZiYuan-3 Multispectral and Stereo Data." *Remote Sensing*, vol. 11, no. 2, 2019, p. 164., doi:10.3390/rs11020164.