Journal of Integrated
SCIENCE & TECHNOLOGY

Semantic segmentation approaches for crop classification with multi-altitude Google Earth imagery

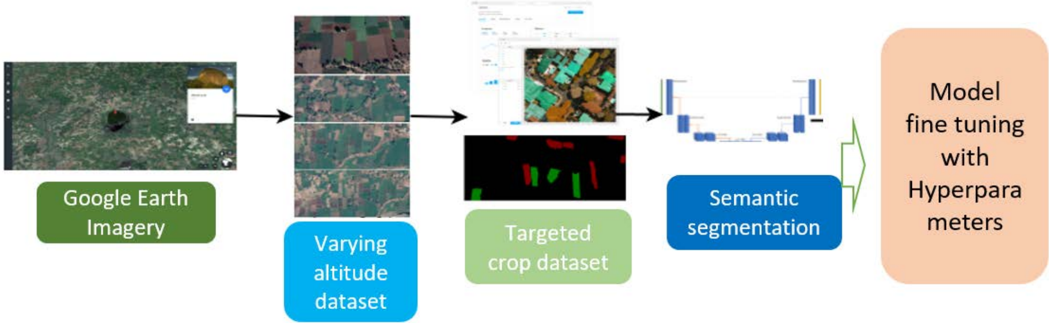
Manan Thakkar[1*], Rakeshkumar Vanzara[2], Ashish Patel[3]

[1]Computer Science and Information Technology, Ganpat University, U. V. Patel College of Engineering, Mehsana, India. [2]Information Technology, Ganpat University, U. V. Patel College of Engineering, Mehsana, India. [3]AI Researcher & Chief Data Scientist, Cygnet Infotech Pvt Ltd, Ahmedabad, India.

Received on: 24-Oct-2023, Accepted and Published on: 11-Jun-2024

Article

ABSTRACT



Based on the advantageous applications of U-Net architecture for semantic segmentation, even with the scarcity of training image samples, herein is reported the application of pixel-wise classification of aerial images and image segmentation with U-Net and U-Net image blocks using google earth imagery. The novel adoption here is to apply segmentation for a single-day varying altitude image dataset collected from a village in Gujarat using the Google Earth's sentinel satellite image views. The classification has been carried out for two major crops, wheat, and Ricinus. The assessment of various architectural frameworks for the segmentation and classification, including ML, DL, and U-Net, by fine-tuning the models with monotonic learning rate (LR) and Cyclic LR, are included in this study. The models are evaluated using standard metrics including accuracy, loss, area under the curve (AUC), sensitivity, and specificity. The U-Net architecture for a dataset with 500m altitudes with a Validation Accuracy of 94.6%, Loss of 3.48%, AUC of 95.76%, Sensitivity of 99.5%, and Specificity of 99.07% and another U-Net image block ResUNet architecture with Cyclic LR for 1000 m altitude outperform the traditional ML, DL algorithms with Validation Accuracy 98.5%, Loss 0.93%, AUC 88.91%, Sensitivity 92.86%, and Specificity 95.42%. Significantly, the U-Net and its image block architectures with Cyclic LR outperform the ML and DL variants.

Keywords: Crop Classification, Remote Sensing, Satellite Imagery, Semantic Segmentation, U-Net

INTRODUCTION

An accurate crop type segmentation can bring information on cultivated crop diversity, crop monitoring, and yield estimation.[1] Manual surveying of the area under cover with the targeted crop and its corresponding yield is an expensive and inaccurate approach.[2,3,4] On the other hand, the last few decades have been considered the era of big data in RS.[5] Several satellites, such as Sentinel A/B, Sentinel 2 A,[6,7] and Landsat-8, were launched in orbit with varying special resolution capabilities. Using these optical and Synthetic Aperture Radar (SAR) satellites, it becomes possible to acquire multi-temporal, multi-altitude, and multispectral data.[8,9] The RS dataset can serve numerous applications, including climate change, urban planning, meteorology, and agriculture.[10,11] An accurate and dynamic way of crop classification in the RS data uses ML, and DL approaches.[12,13] This brings many opportunities in smart farming applications, including crop monitoring, crop yield estimation, precise fertilizer, and pesticide spraying.[14]

Automatic classification of the targeted crop is challenging in many developing nations, especially with small farm holders, due to the special high resolution of RS images. In many research studies, pixel-wise crop classification has been achieved using either a single date image dataset [15] or using multispectral & multi-temporal time series datasets.[16] Over the past decades, ML and DL

[*]Corresponding Author: Manan Thakkar, Research Scholar, Ganpat University, U. V. Patel College of Engineering, Mehsana, India.
Email: mdt01@ganpatuniversity.ac.in

Cite as: *J. Integr. Sci. Technol., 2024, 12(6), 832.*
URN:NBN:sciencein.jist.2024.v12.832
DOI: 10.62110/sciencein.jist.2024.v12.832

architectures have performed well in crop classification and semantic segmentation. Support Vector Machine (SVM) and Random Forest (RF) are the popular ML variant for targeted crop identification using RS image data.[17] In the context of research on ML algorithms, it has been observed that SVM exhibit a significant limitation with respect to their application in big data scenarios that involve a large area to be classified. The reason for this limitation is the high demand for computational resources that SVM algorithms require. Whereas, in the case of RF, multiple features need to be engineered before being forwarded for effective use.[18]

Although ML variants have performed well in classification, overall performance can still be improved for precision agriculture classification. Afterward, DL variants such as Convolutional Neural Networks (CNN) and Long Short-Term Memory (LSTM) have been widely used for crop classification and semantic segmentation.[19] By using RS data for land cover classification and segmentation using single-date images, DL has been observed to work satisfactorily. However, land cover classes such as crop type classification using DL require a multispectral image dataset.[5] With the great success of DL variants, U-Net has been proposed for medical image segmentation. But, U-Net has found its application in many other fields, including agriculture. The U-Net model is suitable for RS image data for classification, and it can better overcome the problems with the imbalanced sample and small sample size.[20] Many researchers have applied U-Net and its image blocks for land cover classification tasks. In the case of land surface classification, features extracted from multiple spectral and textural properties of the surface play a vital role.[21] But, classification of a crop on land is a bit challenging due to changes in land cover properties, such as the shape of the field, size, color, and texture during the growth cycle of the crop throughout the season.

There are several ways to collect RS datasets, including unmanned aerial vehicles (UAV),[22,23,24,25] satellite images, and an open-source platform such as Google Earth.[26] The varying features of multiple satellite spectral bands are due to different spatial and temporal granularity. Existing studies have often used MODIS and Landsat-8 satellite images, with limitations of low revisit and low spatial resolution. Along with RS imagery, high-resolution image acquisition and interpretation lead to high costs and long delays.[27] Using satellite images with multispectral bands in mapping a targeted crop pixel is challenging due to similar spectral properties with other crops in the region. This brings an essential need for a low-cost, open-source, high-resolution dataset to obtain timely and accurate information on crops.[10] Using UAVs with multiple sensors to collect the RGB dataset and Normalized Difference Vegetation Index (NDVI) shows that their classification variant performs better with RGB data than with NDVI data.[22] By keeping this observation into account, instead of relying on the multispectral dataset, the targeted dataset in this paper is the RGB dataset collected using the view available by the sentinel-2 satellite with a good spatial resolution by an input source such as Google Earth. We have collected an image dataset with a single date and varying altitudes of 500m, 600m, 800m, and 1000m for segmenting the image for targeted crops wheat and ricinus.

The critical need for any crop yield estimation is to determine its spatial coverage. To achieve this, the first stage is to classify and determine that particular crop. Due to the multiple spectral information available in the RS dataset of specific satellites, it is possible to recognize real-world objects like vegetation, soil, water, and many more with classification approaches designed based on spectral signatures.[28] There needs to be more than this spectral information to differentiate (classify) the objects[28] (wheat or ricinus). Many existing approaches have used multisource RS data as well. This introduces another challenge of handling different sensor calibration errors and the tedious work of processing multisource data. The collected satellite images are huge, low resolution, and challenging to process using commodity hardware.[29] The need for an open-source image dataset with a good resolution has been identified by keeping these needs for crop classification. Also, to overcome the problem of scarcity[30] of the RS image dataset, efficient segmentation and classification architectures with U-Net and its image blocks have also been introduced in this work.

Research Contributions

The following are the research contributions of this paper:

- We present the working usage of U-Net and its image block architectures for crop pixel classification and semantic segmentation, even with the scarcity of RS image data for training.
- We classify the crop pixels belonging to the targeted crop instead of relying on a land cover classification.
- We evaluate the performance of various existing ML, DL, and U-Net architectures for semantic segmentation by utilizing the single date and RGB picture data rather than depending on multispectral and multitemporal image data. The image data has been captured with Google Earth for the first time to compute the view of the land cover to identify the desired crop.

EXISTING ARCHITECTURES

The usage of CNN variants has proven to be effective for classification and even for segmentation work on various datasets, including satellite images.[31, 32] One big challenge with CNN variants is the need for a sizeable labeled training dataset.[31, 32] The reason behind using U-Net and its variants is the ability of U-Net to achieve precise results in image segmentation, even with few training patches. The U-Net architecture was developed for medical image segmentation and held symmetric paths for down-sampling and up-sampling.[27, 33] The down-sampling part consists of 5 steps, each with 2 convolutions with 3 x 3 kernel size and RELU (Rectified Linear Unit) activation function. This convolution step is followed by max-pooling with 2 x 2 kernel size and dropout. During the down-sampling, at each step, it doubles the size of the feature channel. This down-sampling can be generalized as 5x [2x [[64x2i conv3x3i (previous) + Relu activation], max pool 2x2], dropout] where i in [0..4]. On the other hand, during the up-sampling path, the features get halved.

It starts generating a mask image by concatenating the output of downsampling as an input. This upsampling can be generalized as 5x [[concat (upsample 2x2(previous), output(conv3x3 i)), 2x [64x24 – I x conv3x3 + Relu activation], dropout].

The variations of U-Net architecture to achieve better performance and fast convergence with novel loss functions were designed. These variants have kept the backbone of U-Net, and to make a more deep network for better feature extraction, a few building blocks of U-Net have been replaced or modified with CNN variants blocks. Various possible variants of U-Net implemented in this paper are VGG-UNet, ResNet, Dense-UNet, and UNetPlusPlus. The ResNet is a variation of U-Net that includes multiple layers of residual building blocks.[34] To increase the number of features, input is connected with the convolution layer with kernel size (1, 1). Afterward, it includes residual blocks. In each residual block, three parallel convolutions are used apart from two standard sets of two convolutions from the residual network. It is then followed by the PSPPooling operator.[34] Similarly, other variants VGG-UNet, Dense-UNet, and UNetPlusPlus, can be generated as a modified version of U-Net by adding the CNN variants block of VGG and DenseNet. The comparison between U-Net architecture and its other image block architectures is summarized and compared in Table 1.

Table 1. The Architectural Comparison Between U-NET and Various U-NET Image Blocks

U-Net Architecture	**ResUNet Architecture**	**VGG-UNet Architecture**	**Dense-UNet Architecture**
Down Sampling Path	**Down Sampling Path**	**Down Sampling Path**	**Down Sampling Path**
Conv_2D(3x3) Batch Norm, Activation Fun. } 2	Conv_2D(1x1) Activation Fun.	VGG Block { Conv_2D(3x3) Batch Norm, Activation Fun. } 2	i = input
Max Pool (1x1, stride - 2)	Batch Norm, Activation Fun. Conv_2D(3x3), padding = 'same' } 2 ResBlock	MaxPooling2D	X = { Conv_2D(3x3) Batch Norm, Activation Fun. } 2 - Dense Block
Dropout			x1 = concat[i, x]
Conv_2D(3x3) Batch Norm, Activation Fun. } 2	Conv_2D(1x1) Activation Fun.	VGG Block { Conv_2D(3x3) Batch Norm, Activation Fun. } 2	i = input
Max Pool (1x1, stride - 2)	Batch Norm, Activation Fun. Conv_2D(3x3), padding = 'same' } 2 ResBlock	MaxPooling2D	X = { Conv_2D(3x3) Batch Norm, Activation Fun. } 2 - Dense Block
Dropout			x1 = concat[i, x]
Conv_2D(3x3) Batch Norm, Activation Fun. } 2	Conv_2D(1x1) Activation Fun.	VGG Block { Conv_2D(3x3) Batch Norm, Activation Fun. } 3	i = input
Max Pool (1x1, stride - 2)	Batch Norm, Activation Fun. Conv_2D(3x3), padding = 'same' } 2 ResBlock	MaxPooling2D	X = { Conv_2D(3x3) Batch Norm, Activation Fun. } 2 - Dense Block
Dropout			x1 = concat[i, x]
Conv_2D(3x3) Batch Norm, Activation Fun. } 2	Conv_2D(1x1) Activation Fun.	VGG Block { Conv_2D(3x3) Batch Norm, Activation Fun. } 3	i = input
Max Pool (1x1, stride - 2)	Batch Norm, Activation Fun. Conv_2D(3x3), padding = 'same' } 2 ResBlock	MaxPooling2D	X = { Conv_2D(3x3) Batch Norm, Activation Fun. } 2 - Dense Block
Dropout			x1 = concat[i, x]
Conv_2D(3x3) Batch Norm, Activation Fun. } 2	Conv_2D(1x1) Activation Fun.	VGG Block { Conv_2D(3x3) Batch Norm, Activation Fun. } 3	i = input
Max Pool (1x1, stride - 2)	Batch Norm, Activation Fun. Conv_2D(3x3), padding = 'same' } 2 ResBlock	MaxPooling2D	X = { Conv_2D(3x3) Batch Norm, Activation Fun. } 2 - Dense Block
Dropout			x1 = concat[i, x]
	PSPPooling		
Up Sampling Path	**Up Sampling Path**	**Up Sampling Path**	**Up Sampling Path**
Conv2DTranspose(3x3) Concatenate, Dropout	Conv_2D(1x1) UpSampling2D,	Conv2DTranspose(2x2) BatchNormalization, Activation	UpSampling2D

Conv_2D(3x3)	Activation Fun. Concatenate Combine Conv2D(1x1)	Concatenate	concatenate
	Batch Norm, Activation Fun. 2 ResBlock Conv_2D(3x3), padding = ‘same’	Conv2D(3x3) BatchNormalization, 2 Activation	Conv2D(3x3) BatchNormalization, 2 Activation
Conv2DTranspose(3x3) Concatenate, Dropout	Conv_2D(1x1) UpSampling2D,	Conv2DTranspose(2x2) BatchNormalization, Activation	UpSampling2D
Conv_2D(3x3)	Activation Fun. Concatenate Combine Conv2D(1x1)	Concatenate	concatenate
	Batch Norm, Activation Fun. 2 ResBlock Conv_2D(3x3), padding = ‘same’	Conv2D(3x3) BatchNormalization, 2 Activation	Conv2D(3x3) BatchNormalization, 2 Activation
Conv2DTranspose(3x3) Concatenate, Dropout	Conv_2D(1x1) UpSampling2D,	Conv2DTranspose(2x2) BatchNormalization, Activation	UpSampling2D
Conv_2D(3x3)	Activation Fun. Concatenate Combine Conv2D(1x1)	Concatenate	Concatenate
	Batch Norm, Activation Fun. 2 ResBlock Conv_2D(3x3), padding = ‘same’	Conv2D(3x3) BatchNormalization, 2 Activation	Conv2D(3x3) BatchNormalization, 2 Activation
Conv2DTranspose(3x3) Concatenate, Dropout	Conv_2D(1x1) UpSampling2D,	Conv2DTranspose(2x2) BatchNormalization, Activation	UpSampling2D
Conv_2D(3x3)	Activation Fun. Concatenate Combine Conv2D(1x1)	Concatenate	Concatenate
	Batch Norm, Activation Fun. 2 ResBlock Conv_2D(3x3), padding = ‘same’	Conv2D(3x3) BatchNormalization, 2 Activation	Conv2D(3x3) BatchNormalization, 2 Activation
Conv2DTranspose(3x3) Concatenate, Dropout	Conv_2D(1x1) UpSampling2D,	Conv2D(3x3) BatchNormalization, 2 Activation	Conv2D Activation – relu 3
Conv_2D(3x3)	Activation Fun. Concatenate Combine Conv2D(1x1)		Conv2D Activation – sigmoid
	Batch Norm, Activation Fun. 2 ResBlock Conv_2D(3x3), padding = ‘same’		
	Activation Fun. Concatenate Combine Conv2D(1x1) PSPPooling		

METHODOLOGY

Data Collection

To avoid the challenge of collecting an image covering a vast area using sentinel-1 or other satellites and then segmenting that image for better labeling and processing, an open-source dataset from Google Earth with Sentinel satellite's RGB views has been taken as a source of the dataset. Wheat and Ricinus are the main winter-season crops in North Gujarat, India region. To generate the dataset with its patches, 21st November 2018, the date selected from Google Earth pro views. This day fall into early crop growth stage of winter seasoned crop. For this date, the image dataset of the entire Fatepura village of Mahesana district in Gujarat state is chosen. The study area to collect this dataset is highlighted in Figure 1.

Dataset Description

The dataset has been collected using Google Earth and Smart GIS with diversified altitudes such as 500m, 600m, 800m, and 1000m. A total of 288 images were collected with 80, 80, 80, and 48 image samples with altitudes of 500m, 600m, 800m, and 1000m, respectively. Out of these images, 64 images with 500m, 55 images with 600m, 61 images of 800m, and 34 images with 1000m altitude having wheat and Ricinus crop patches are labeled and used further. The Sentinel-2 imagery dataset collected using the open-source platform Google Earth comprises two categories: for training, it takes a three-channel color image, and for the mask, it takes one channel grayscale image. The summary of the image dataset is mentioned below in Table 2.

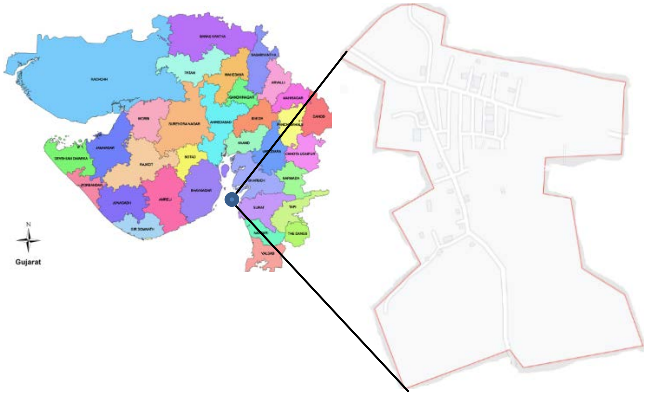


Figure 1. Sample Remote Sensing Image Data for Surface Reflectance, Temperature

Table 2. Count of Raw Images and Labelled Images Patches in Dataset

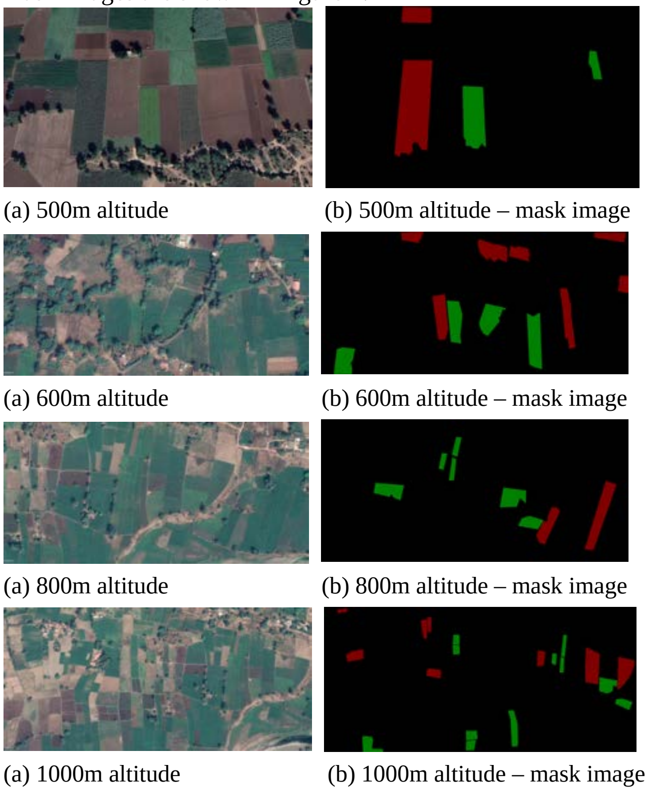
Altitude	Raw Dataset	Labeled Dataset	Crop wise Patches for Training for each altitude
500m	80	64	Wheat – 110 Ricinus - 154
600m	80	55	
800m	80	61	
1000m	48	34	
Total	288 images	214 images	1056 Patches

Data Labeling and mask Image generation

The labeling of actual collected varying altitude Google Earth images has been carried out using the labelme graphical annotation tool.[35] After labeling, the JSON formatted file for each labeled image gets generated. We have used the below-mentioned command to convert the JSON formatted file into a mask image and segmented dataset. It generates the standard files as an actual image in PNG format, uint8 label file, visualization of an actual PNG file, and txt file with label names from JSON file.[35]

```
labelme_json_to_dataset sample.json -o sample_json
```

By leveraging the knowledge and insights of the local community, it become possible to create a more accurate and detailed labeling for Google Earth images with patches of wheat and ricinus crops in the targeted area. The sample image dataset of varying altitudes (500m, 600m, 800m, and 1000m) along with its mask images are shown in Figure 2.



(a) 500m altitude (b) 500m altitude – mask image

(a) 600m altitude (b) 600m altitude – mask image

(a) 800m altitude (b) 800m altitude – mask image

(a) 1000m altitude (b) 1000m altitude – mask image

Figure 2. Sample Google Earth images of varying altitudes and their mask image dataset (Legend in mask image: Red color – Ricinus, Green color – Wheat)

Splitting the Dataset

To train the model and assess the performance of all the proposed model architectures, the actual image dataset and its corresponding mask images are divided into two parts – training and validation. Each altitude dataset is divided into 90% for training purposes and the remaining 10% to validate the model.

Selection of Image Segmentation and Classification Variant – Frameworks

ML approaches used by the authors in[31, 32] achieved promising results in classification and segmentation. Due to manual feature extraction and the need for a huge dataset for training, these approaches lead to a huge resource and time consuming for fine-tuning the model. To avoid this manual feature extraction need, a hybrid variant has been proposed with deep feature extraction using a DL variant (e.g., VGG) and passing these extracted features to the ML variants to fine-tune the model. The model and its performance with ML variants and hybrid approach are taken as a benchmark to compare and analyze the performance. CNN variants require a large amount of labeled dataset to fine-tune the model [31], which leads to generating a model with U-Net and its image blocks [28, 31] which requires less dataset to fine-tune the model.

Due to promising results of U-Net and other image block architectures in various applications of RS datasets, U-Net[33], ResUNet[34], Dense-UNet[36], VGG-UNet[37], and UnetPlusPlus[38] are implemented on our dataset. Hyperparameter tuning has been carried out to improve the performance of U-Net image block architectures. During the training, there is critical importance to the learning rate parameter. The intention behind it is to tune the model quickly with the best learning rate to improve the model performance. There are two types of learning rates: Monotonic learning rate (LR) and Cyclic LR with lower bound LR, upper bound LR, and cyclic learning rate policy. Thus, three versions of each U-Net variant are implemented, including the actual variant, the variant with LR, and the variant with Cyclic LR.

Figures 3, 4, 5, and 6 show all possible frameworks by keeping different possible algorithms to perform semantic segmentation & crop classification. Figure 3, Figure 4, and Figure 5 & Figure 6 are crop classification architectures with U-Net, ML variants, and hybrid approaches, respectively.

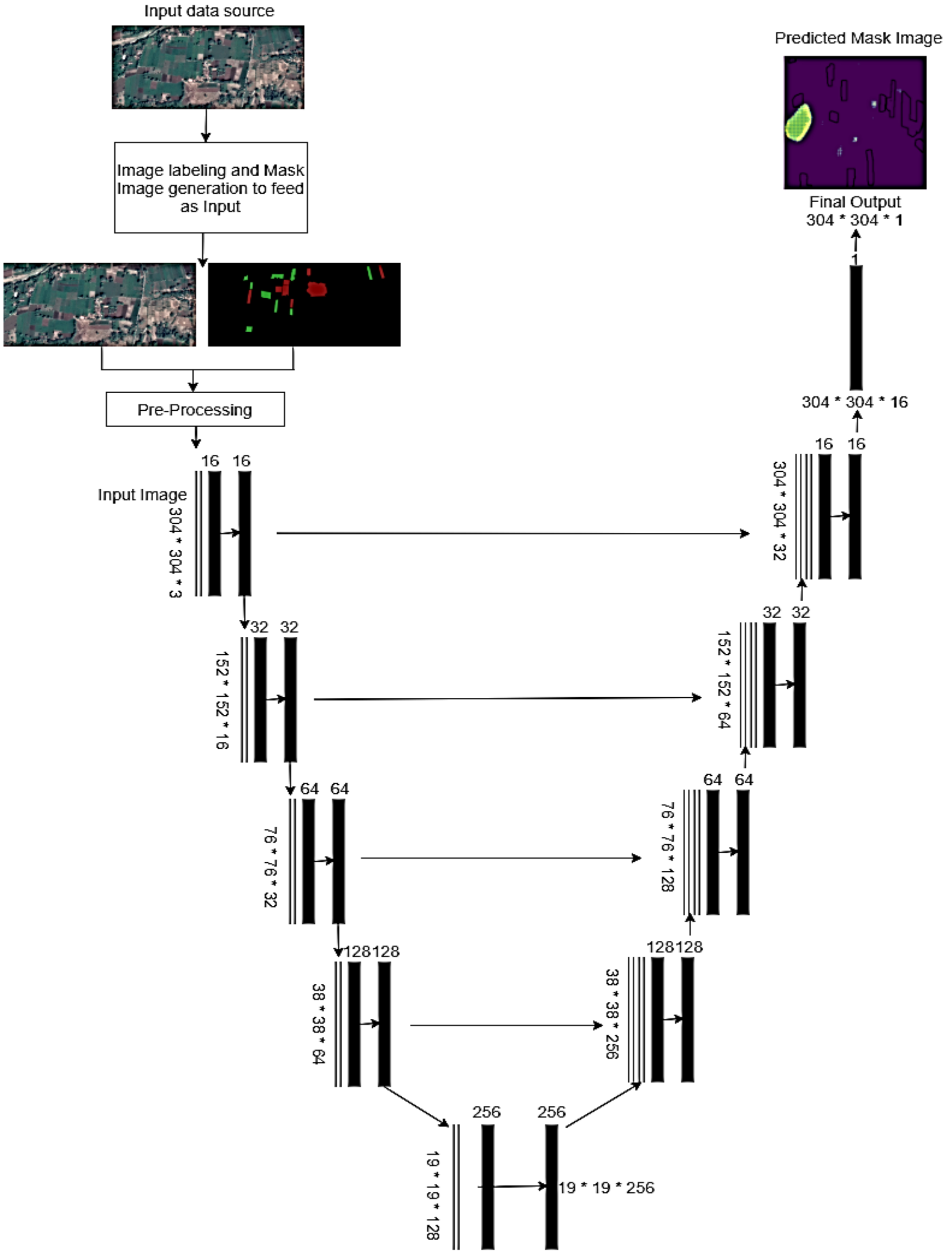


Figure 3. Crop Classification using U-Net and various U-Net Image Blocks

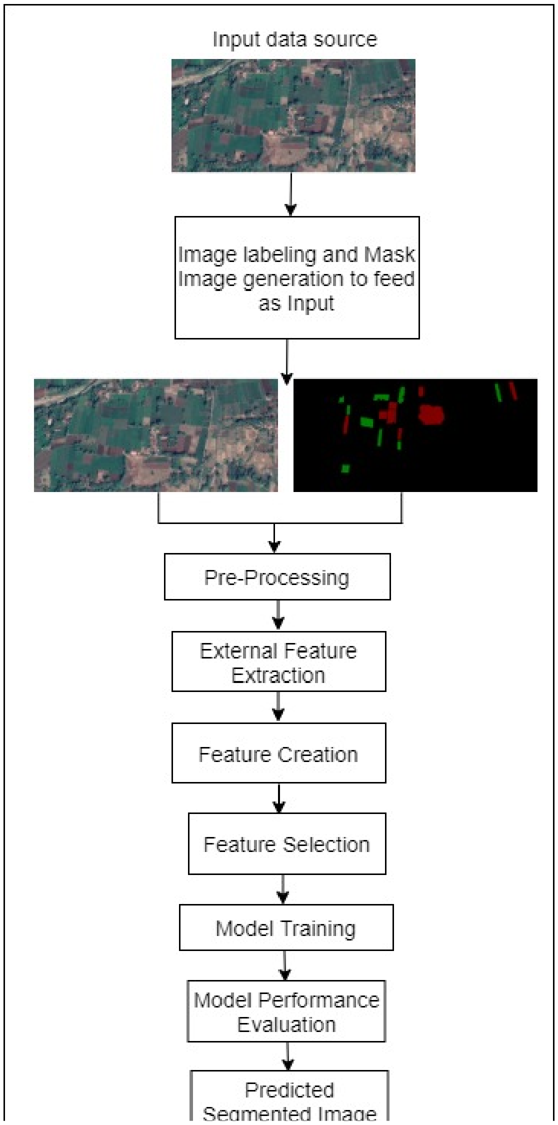

Input data source
Image labeling and Mask Image generation to feed as Input
Pre-Processing
External Feature Extraction
Feature Creation
Feature Selection
Model Training
Model Performance Evaluation
Predicted Segmented Image

Figure 4. Framework for Semantic Segmentation & Crop Classification using Machine Learning

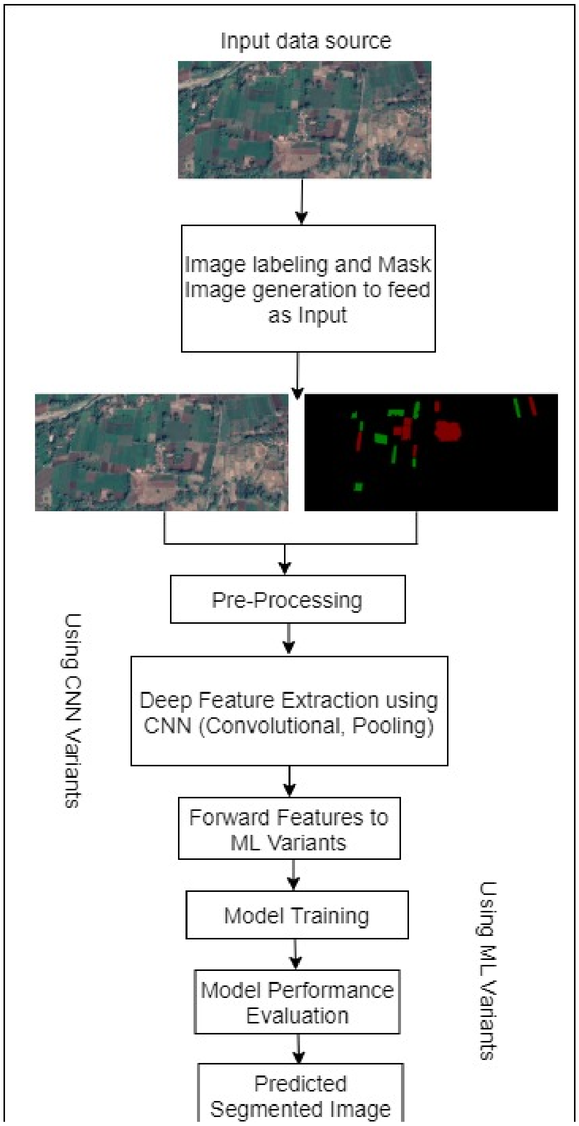

Input data source
Image labeling and Mask Image generation to feed as Input
Using CNN Variants
Pre-Processing
Deep Feature Extraction using CNN (Convolutional, Pooling)
Forward Features to ML Variants
Model Training
Using ML Variants
Model Performance Evaluation
Predicted Segmented Image

Figure 5. Framework for Semantic Segmentation & Crop Classification using ML-DL Hybrid Approach

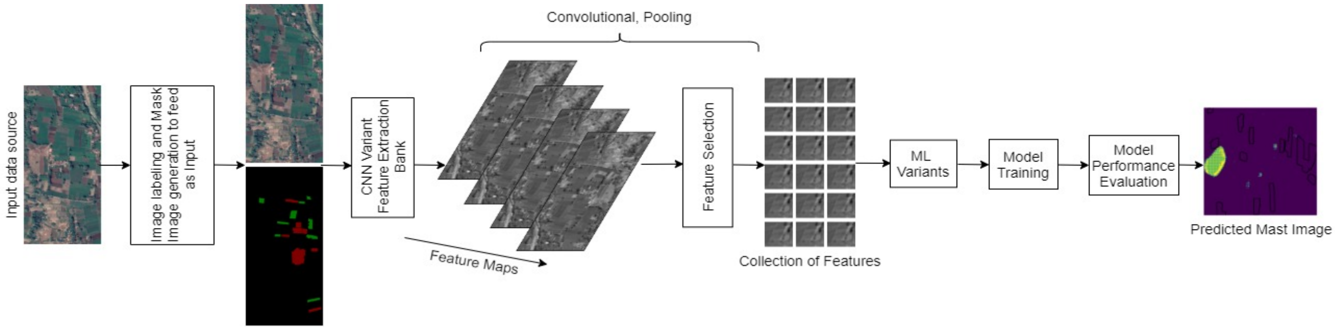


Figure 6. Overall flow for Semantic Segmentation & Crop Classification using Hybrid Approach

Experimental Results and Analysis

Simulation Setup

Experiments were carried out over Google Colab Pro with GPU specifications as 1xTesla K80 (4692 cores), P100 (3584 cores), T4 (2560 Tensor Cores), and 32GB GDDR5 vRAM with 2 x vCPU. The U-Net and its image block variants have been implemented by setting up the hyperparameters mentioned in Table 3.

Table 3. Fine-Tuned Parameters in Implementataion

Parameter	Value
Optimizer	Adam
Loss function	binary_crossentropy
Kernel_initializer	he_normal
Epochs	200
Batch Size	6
Dropout	0.05
Batch Normalization	True
For LR Scheduler	
Decay_rate	0.1
Decay_step	10
For Cyclic LR Scheduler	
Base_lr	0.0001
Max_lr	0.000001
Mode	Triangular

Evaluation Metrics

The performance of proposed models for semantic segmentation and classification has been evaluated by keeping a test size of 10% of the available dataset. The performance of the models has been evaluated on the metrics of Accuracy, Precision, F1-Score, AUC, Sensitivity, and Specificity.

Table 4. Performance Measurement Metrics [39, 40]

Metrics	Formula	
Accuracy	$Accuracy = \frac{TP + TN}{(TP + FP + TN + FN)}$	
Precision	$Precision = \frac{TP}{(TP + FP)}$	
F1-Score	$F1 - Score = \frac{2 * Precision * Recall}{(Precision + Recall)}$	
AUC	**AUC Range** 0.9 < AUC < 1.0 0.8 < AUC < 0.9 0.7 < AUC < 0.8 0.6 < AUC < 0.7	**Classification** Excellent Good Worthless Not Good
Sensitivity OR Recall	$Sensitivity = \frac{TP}{(TP + FN)}$	
Specificity	$Specificity = \frac{TN}{(TN + FP)}$	

Result Analysis

Initially, the evaluation was carried out using the traditional ML and hybrid variants - RF, SVM, VGG + RF, and VGG + SVM. The performance of these models has been taken as a benchmark. The performance of these variants has been evaluated by considering accuracy, precision, recall, and F1-Score metrics. After experimental analysis using U-Net, U-Net Image blocks, ML variants, and hybrid approaches on the performance measurement metrics, U-Net, and U-Net Image blocks outperform the traditional. The highest accuracy achieved using the traditional approach is 85.9% with a 500m altitude dataset and that using a hybrid approach VGG + SVM. This means the model can classify only 85.9% of the pixels of an image correctly. Followed by that, using ML variant RF, 81.2% accuracy was achieved at an altitude of 500m. Again there is a lot of room for improvement as the model is capable of correctly classifying 81.2% pixels of an image.

Using U-Net and U-Net Image blocks, the highest 98.5% accuracy at an altitude of 1000m and 94.6% accuracy at an altitude of 500m has been observed. As U-Net and U-Net Image block architectures outperform the traditional approaches for semantic segmentation, the assessment of different U-Net Image blocks has been carried out. It has been observed that increasing the altitude decreases the resolution of an image. Due to that model is unable to extract fine-grained details of crop features. This result leads to a lowering of the overall model performance.

To assess the performance of U-Net and U-Net Image blocks, the experiments have been carried out for a single date, single altitude image dataset, and single date multi altitude dataset. The evaluation has been carried out using U-Net, ResUNet, VGG-UNet, Dense-UNet, and UNetPlusPlus. By hyperparameter tuning, two more versions of all five variants have been implemented, i.e., a variant with LR and CyclicLR parameter tuning. Each experiment has been implemented for 200 epochs. In the case of CyclicLR, base_lr = 0.0001, max_lr =0.000001, and triangular mode selected.

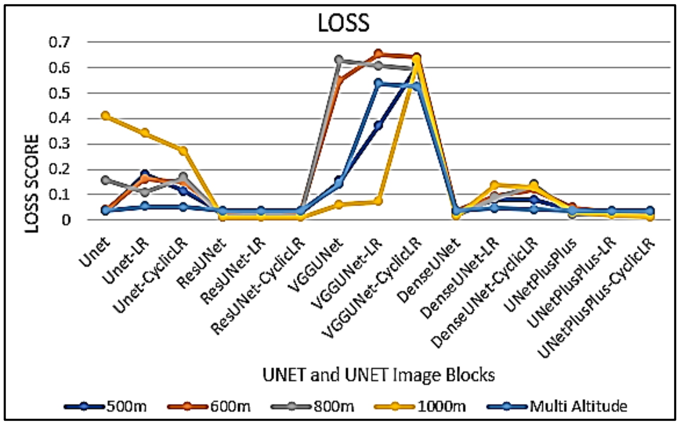


Figure 7. Validation Loss comparison of UNET and UNET Image Block Architectures, with Varying Altitude Image Dataset

Table 5. Performance Evaluation Using Traditional Approaches for Crop Classification

	ACCURACY				PRECISION				RECALL				F1-SCORE			
Altitude	RF	SVM	VGG + RF	VGG + SVM	RF	SVM	VGG + RF	VGG + SVM	RF	SVM	VGG + RF	VGG + SVM	RF	SVM	VGG + RF	VGG + SVM
500m	**0.812**	0.805	0.771	**0.859**	**0.77**	0.65	0.67	**0.74**	**0.81**	0.81	0.77	**0.86**	**0.75**	0.72	0.71	**0.8**
600m	0.636	0.563	0.583	0.346	0.62	0.51	0.55	0.64	0.64	0.56	0.58	0.35	0.61	0.51	0.56	0.18
800m	0.600	0.483	0.503	0.483	0.59	0.4	0.49	0.46	0.6	0.48	0.5	0.48	0.59	0.32	0.5	0.33
1000m	0.627	0.427	0.511	0.446	0.61	0.18	0.49	0.53	0.63	0.43	0.51	0.45	0.61	0.26	0.49	0.29

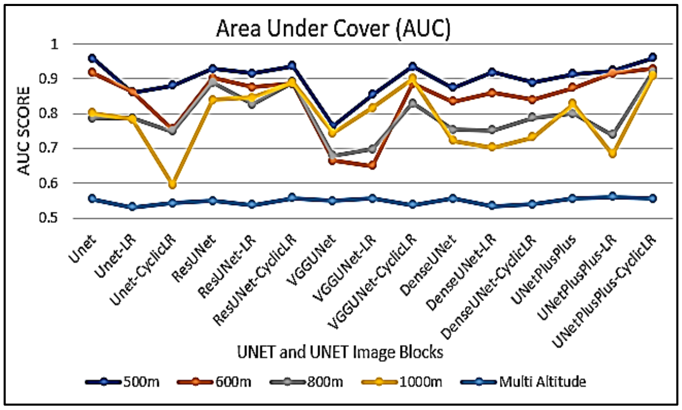


Figure 8. Validation AUC Score comparison of UNET and UNET Image Block Architectures with Varying Altitude Image Dataset

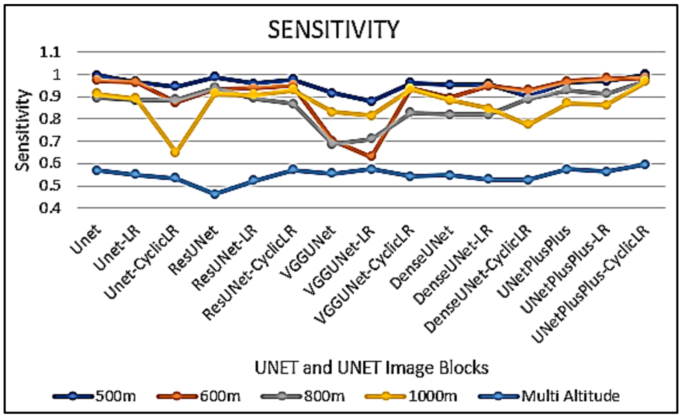


Figure 9. Sensitivity comparison of UNET and UNET Image Block Architectures with Varying Altitude Image Dataset

The performance analysis using the validation loss score has varied between 0.009305 to 0.65213 values. It has been observed that the lowest loss was achieved in the 1000m altitude dataset. This proves that the higher the altitude, the better the features of the satellite images. Based on the performance analysis on validation loss, ResUNet outperforms the other U-Net Image Block architectures because of its Res blocks to extract more features in the down-sampling stage of U-Net. The performance analysis on the parameter AUC score has the lowest value in the multi-altitude dataset because of different image feature importance. Our experiments with a single altitude image dataset observed scales between 0.65 and 0.95 with VGG-U-NET with LR and UNetPlusPlus with CyclicLR, respectively. The variants UNET, ResUNet, and UNetPlusPlus outperform with AUC scores of 0.95, 0.92, and 0.91, respectively (Excellent class as per AUC score range) with a 500m altitude dataset.

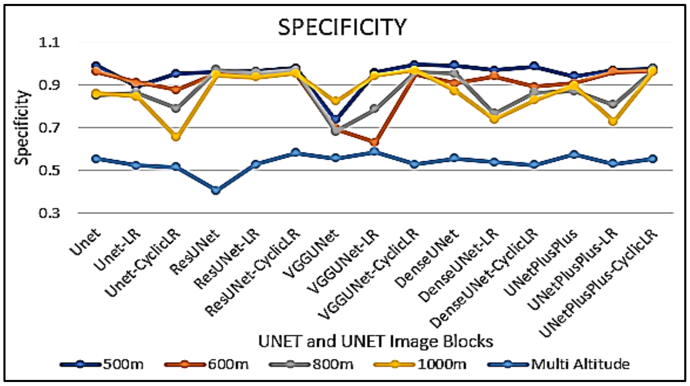
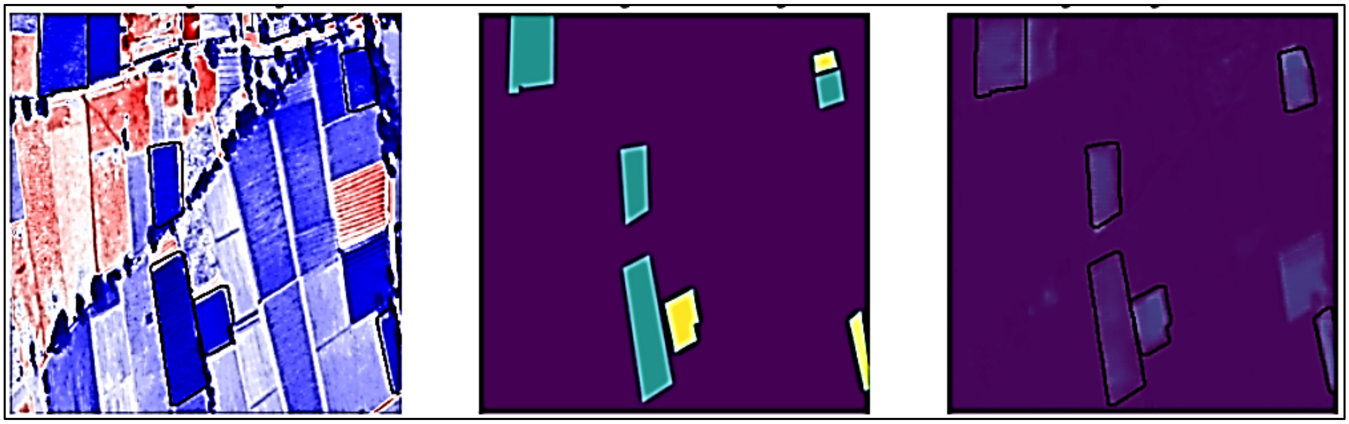


Figure 10. Specificity comparison of UNET and UNET Image Block Architectures with Varying Altitude Image Dataset

The performance analysis on metrics Sensitivity and Specificity are shown in Figures 9 and 10, respectively. The Sensitivity score is in the range of 0.6299 and 0.9977 with VGG-UNet with LR and UNetPlusPlus with CyclicLR, respectively. On the other hand, the specificity has values in the range of 0.6307 and 0.9775 with VGG-UNet with LR and UNetPlusPlus with CyclicLR, respectively. Again, the multi-altitude dataset performs poorly because of improper weight parameter distribution. Based on the result analysis, UNetPlusPlus and ResUNet with CyclicLR outperform the other variants. Tables 6 - 13 and Figures 11 - 18 provide an extensive empirical evaluation of the best five results, using performance metrics with different segmentation architectures- U-Net, ResUNet, U-Net with CyclicLR, and ResUNet with CyclicLR. The analysis is depicted for the dataset with 500m and 1000m altitudes. Figures 11 - 18 indicates the actual satellite image, mask image, and satellite-predicted image from left to right. The actual satellite images from Figures 11 - 18 are derived using cmap from the sample images shown in Figure 2. The image's color is visible due to the seismic cmap value. Even the predicted satellite image is a grayscale image with the same cmap.

Table 6. Top 5 Results On Performance Metrics Using U-NET (500M Altitute)

VAL_ACCURACY	VAL_LOSS	VAL_AUC	VAL_SENSITIVITY	VAL_SPECIFICITY
0.946383715	0.034822	0.9576	0.995069921	0.990714252
0.945816457	0.034863	0.95744	0.994781613	0.988773763
0.942497611	0.034863	0.95658	0.994723916	0.987774134
0.930182159	0.034866	0.95532	0.99463743	0.987596095
0.798380613	0.034866	0.95511	0.99463743	0.987422943

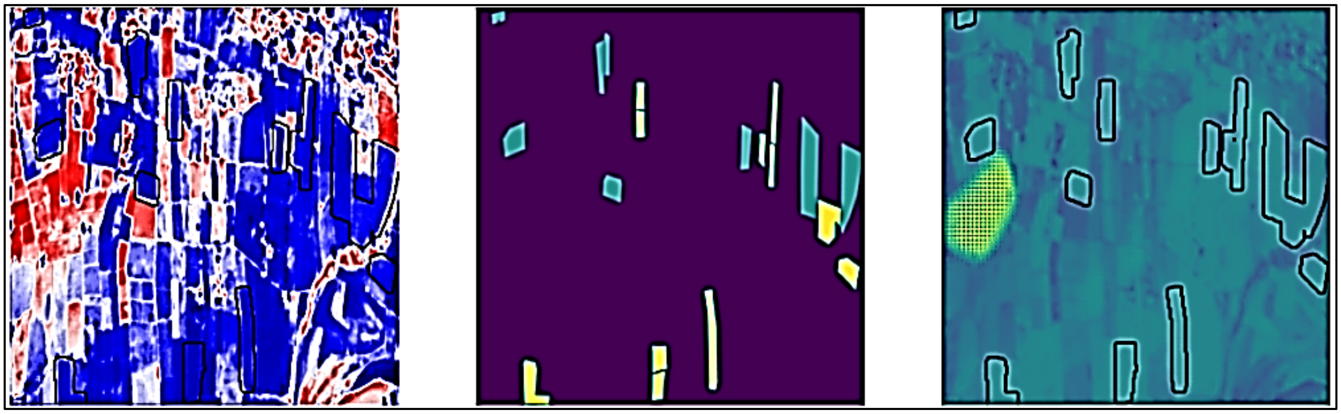


Satellite Image **Satellite Mask Image** **Satellite Image Predicted**

Figure 11. Predicted satellite image result using U-Net with Google earth image at 500m altitudes (left to right – Actual satellite image, Mask image, and Predicted Image)

Table 7. Top 5 Results On Performance Metrics Using U-NET (1000M Altitude)

VAL_ACCURACY	VAL_LOSS	VAL_AUC	VAL_SENSITIVITY	VAL_SPECIFICITY
0.985292077	0.407958	0.80187	0.912451744	0.863104582
0.985289335	0.408197	0.79489	0.909876764	0.860317886
0.985286653	0.408525	0.78804	0.908957124	0.823988914
0.985283911	0.408711	0.78169	0.905462563	0.821666181
0.985254169	0.408898	0.77696	0.901784062	0.816869676

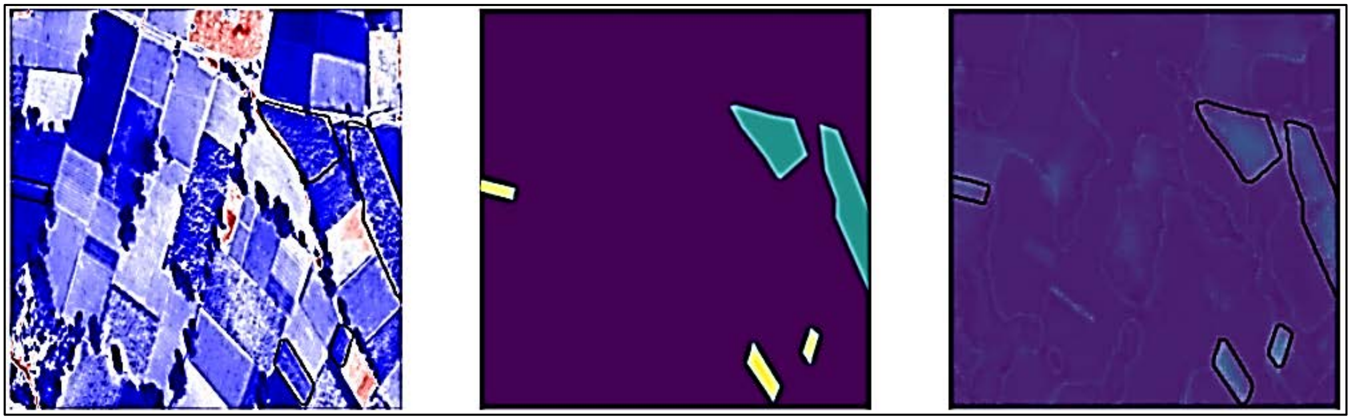


Satellite Image **Satellite Mask Image** **Satellite Image Predicted**

Figure 12. Predicted satellite image result using U-Net with Google Earth Image at 1000m altitude (left to right – Actual satellite image, Mask image, and Predicted Image)

Table 8. Top 5 Results On Performance Metrics Using U-NET With CyclicLR (500M Altitude)

VAL_ACCURACY	VAL_LOSS	VAL_AUC	VAL_SENSITIVITY	VAL_SPECIFICITY
0.946383715	0.1127313	0.880612075	0.944442868	0.953723073
0.946382225	0.112750292	0.875685275	0.941357911	0.950637579
0.946380675	0.112821661	0.875677347	0.940320015	0.947787344
0.946379125	0.113185838	0.875555575	0.940031707	0.946666837
0.946377575	0.113223553	0.875421762	0.939801037	0.945649266

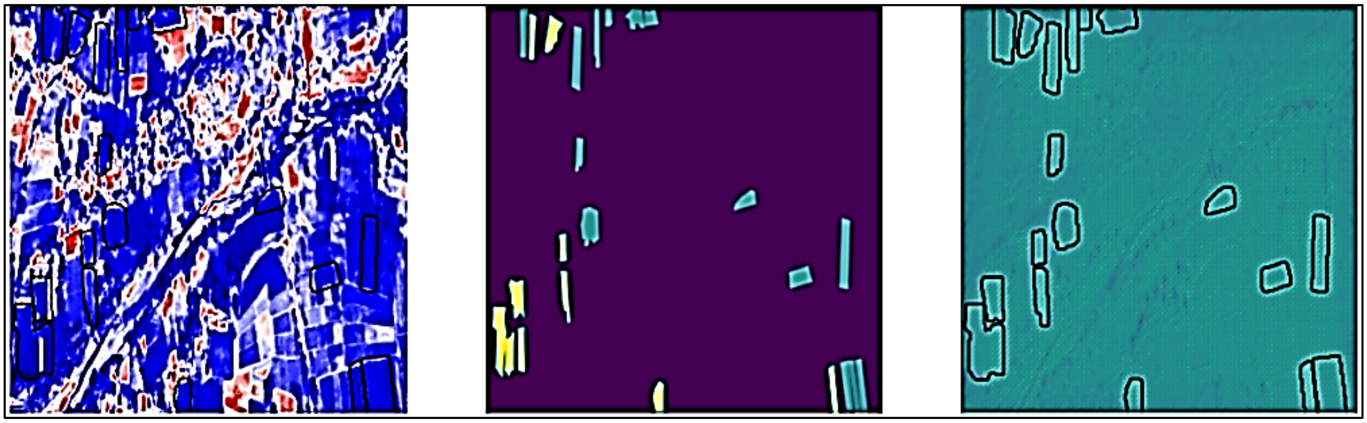


Satellite Image **Satellite Mask Image** **Satellite Image Predicted**

Figure 13. Predicted satellite image result using U-Net with CyclicLR for Google earth image at 500m altitudes (left to right – Actual satellite image, Mask image, and Predicted Image)

Table 9. Top 5 Results on Performance Metrics Using U-NET With CyclicLR (1000M Altitude)

VAL_ACCURACY	VAL_LOSS	VAL_AUC	VAL_SENSITIVITY	VAL_SPECIFICITY
0.985292077	0.270254165	0.594528675	0.649439037	0.65341121
0.985289335	0.270484596	0.591830432	0.632885754	0.633371472
0.985286653	0.270520389	0.591815054	0.627368033	0.628424048
0.985283911	0.270576715	0.591296911	0.627184093	0.619687736
0.985281229	0.270916969	0.590423822	0.626632333	0.613867164

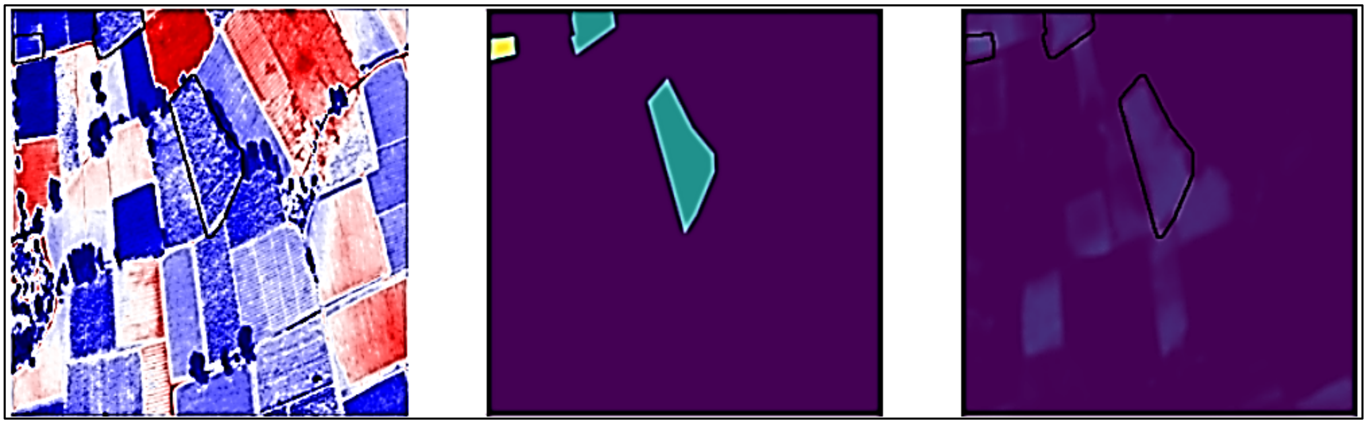


Satellite Image **Satellite Mask Image** **Satellite Image Predicted**

Figure 14. Predicted satellite image result using U-Net with CyclicLR for Google earth image at 1000m altitude (left to right – Actual satellite image, Mask image, and Predicted Image)

Table 10. Top 5 Results On Performance Metrics Using RESUNET (500M Altitude)

VAL_ACCURACY	VAL_LOSS	VAL_AUC	VAL_SENSITIVITY	VAL_SPECIFICITY
0.946383715	0.029179	0.92817	0.98714143	0.962334216
0.946383715	0.029237	0.92643	0.9609918	0.96079886
0.946383715	0.02924	0.90786	0.943174303	0.960754752
0.946383715	0.029241	0.90424	0.930719316	0.960625708
0.946383715	0.029247	0.90344	0.928931832	0.96060282

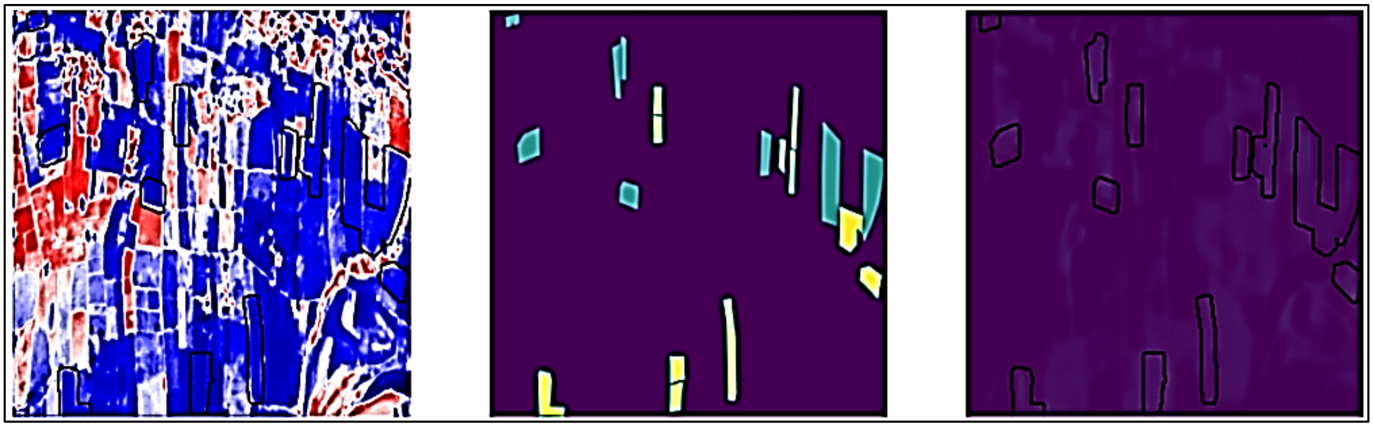


Satellite Image **Satellite Mask Image** **Satellite Image Predicted**

Figure 15. Predicted satellite image result using ResUNet with Google earth image at 500m altitudes (left to right – Actual satellite image, Mask image, and Predicted Image)

Table 11. Top 5 Results on Performance Metrics Using RESUNET (1000M Altitude)

VAL_ACCURACY	VAL_LOSS	VAL_AUC	VAL_SENSITIVITY	VAL_SPECIFICITY
0.985292077	0.009902	0.83993	0.913923144	0.945295095
0.985292077	0.010116	0.83986	0.904910803	0.941418409
0.985292077	0.010139	0.83674	0.875666738	0.939891875
0.985292077	0.010141	0.83649	0.861872375	0.935976744
0.985292077	0.010249	0.83336	0.83299613	0.935938299

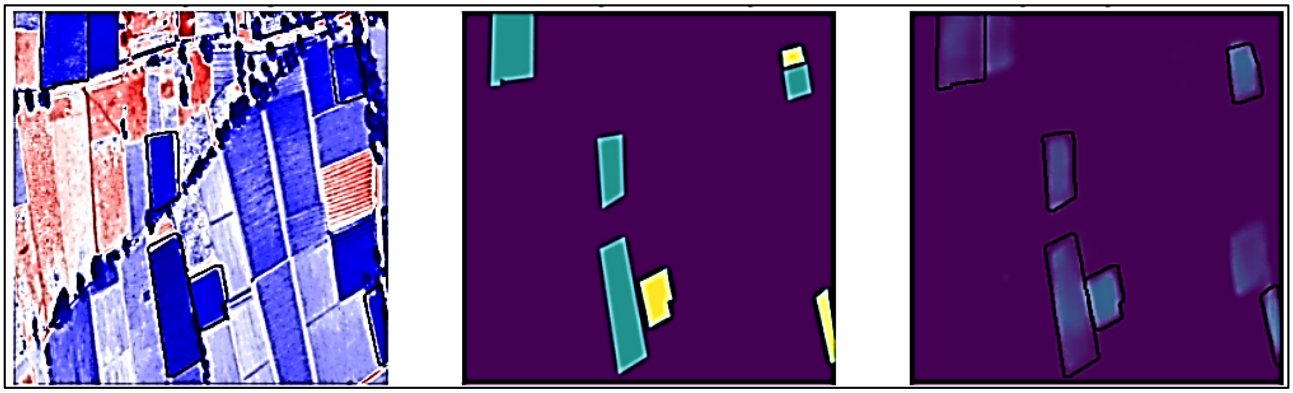


Satellite Image **Satellite Mask Image** **Satellite Image Predicted**

Figure 16. Predicted satellite image result using ResUNET with Google earth image at 1000m altitude (left to right – Actual satellite image, Mask image, and Predicted Image)

Table 12. Top 5 Results On Performance Metrics Using RESUNET With CyclicLR (500M Altitude)

VAL_ACCURACY	**VAL_LOSS**	**VAL_AUC**	**VAL_SENSITIVITY**	**VAL_SPECIFICITY**
0.946383715	0.02822496	0.936778605	0.977165937	0.9795174
0.946383715	0.028389653	0.929304302	0.964768648	0.979159713
0.946383715	0.028770709	0.922254741	0.954821944	0.979029
0.946383715	0.029098403	0.920098066	0.946172714	0.977975488
0.946383715	0.029218972	0.920069396	0.944298685	0.977782726

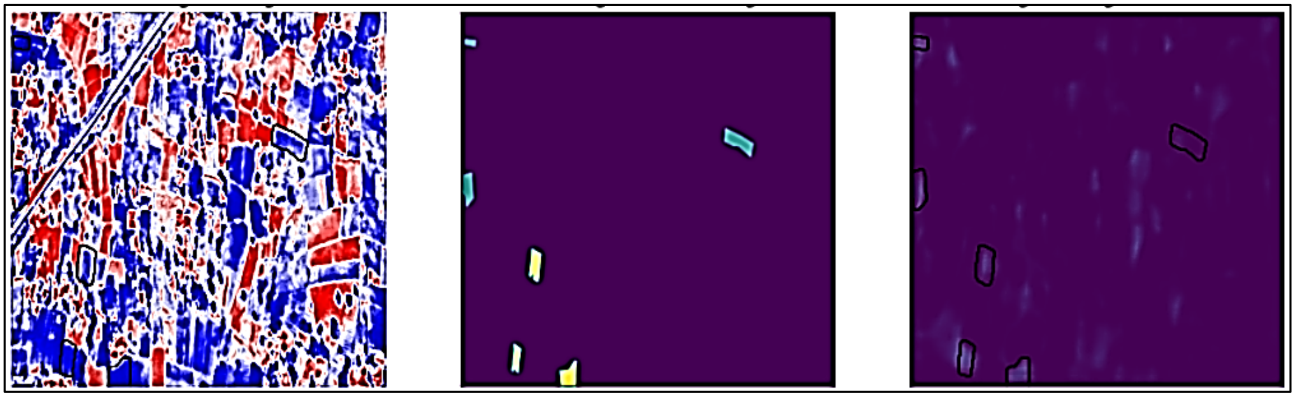


Satellite Image **Satellite Mask Image** **Satellite Image Predicted**

Figure 17. Predicted satellite image result using ResUNet with CyclicLR for Google earth image at 500m altitudes (left to right – Actual satellite image, Mask image, and Predicted Image)

Table 14. Top 5 Results On Performance Metrics Using RESUNET With CyclicLR (1000M Altitude)

VAL_ACCURACY	**VAL_LOSS**	**VAL_AUC**	**VAL_SENSITIVITY**	**VAL_SPECIFICITY**
0.985292077	**0.009305747**	0.889189422	0.928637087	0.954209864
0.985292077	0.009388614	0.881810009	0.928085327	0.948457956
0.985292077	0.009436878	0.868653774	0.926981807	0.946148992
0.985292077	0.009484857	0.868309617	0.899209142	0.945588887
0.985292077	0.009546737	0.863807619	0.891300321	0.945289612



Satellite Image **Satellite Mask Image** **Satellite Image Predicted**

Figure 18. Predicted satellite image result using ResUNet with CyclicLR for Google Earth Image at 1000m altitude (left to right – Actual satellite image, Mask image, and Predicted Image)

Key Takeaways

The following key takeaways have been identified as a result of the novel source of RS data collection and assessment of all possible methods for crop classification and semantic segmentation:

- The effective use of an RGB image dataset with a single date and varying altitudes collected from an open-source Google Earth.
- Crop classification and semantic segmentation ability by identifying the region of interest using U-Net and U-Net image blocks in the face of data scarcity.
- Impact of different LR schedulers on overall performance by fine-tuning the model.
- Observations after assessments with ML, DL, hybrid ML+DL, and U-Net variants for crop classification can open a new roadmap for many researchers.

Conclusion

An extensive experimental study and analysis with the top-5 performance of the model over varying altitude datasets have been represented in this work. We observed that with hyperparameter tuning using CyclicLR, the ResUNet and UNetPlusPlus models outperform the other models evaluated. As per the empirical analysis with multi altitude single date dataset, it can be observed that altitude plays a major role in the overall performance and also shows possibilities to achieve the best crop classification and segmentation without a multi-spectral dataset. As per the analysis, 1000m altitude is best suitable to precisely achieve the crop classification. The best overall performance achieved with metrics such as accuracy of 98.52%, loss of 0.9%, AUC of 0.88, Sensitivity 0.92, and Specificity of 0.95 has been achieved using ResUNet with CyclicLR on a 1000m altitude dataset. The high-resolution crop classification and segmentation demonstrated in this work has significant potential for applications in precision agriculture, land-use planning, and vegetation monitoring. By enabling accurate crop differentiation, our models can support optimized resource allocation and data-driven decisions to increase farm productivity and profitability over time.

The main objective of this research work is to perform a comparative analysis between U-Net and U-Net image blocks for crop segmentation. Additionally, as a further comparative assessment, it is necessary to evaluate the performance of other advanced techniques such as DeepLab, TranUNet, and SegFormer to achieve even higher accuracy and efficiency in crop segmentation. To further expand on the findings of this paper, future research can be conducted using a multi-temporal dataset that covers a wider range of crops and a larger geographical area such as a state or country.

Abbreviations used:

Abbreviations	Meanings
AUC	Area Under Cover
CNN	Convolutional Neural Network
Conv2D	2D Convolution Layer
Dense-UNet	Dense Structure with U-Net
DL	Deep Learning
ENN	Ensemble Neural Network
FN	False Negative
FP	False Positive
LR	Learning Rate
LSTM	Long Short-Term Memory
NDVI	Normalized Difference Vegetation Index
ML	Machine Learning
RELU	Rectified Linear Unit
ResUNet	Deep Residual U-Net
RF	Random Forest
RS	Remote Sensing
SAR	Synthetic Aperture Radar
SVM	Support Vector Machine
TN	True Negative
TP	True Positive
UAV	Unmanned Aerial Vehicle
VGG	Visual Geometry Group

Conflict of Interest Statement

The authors declare that there is no conflict of interest for this work.

References

1. R. Rustowicz, R. Cheong, L. Wang, et al. Semantic segmentation of crop type in Africa: A novel dataset and analysis of deep learning methods. *IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit. Work.* **2019**, 2019-June, 75–82.
2. T. van Klompenburg, A. Kassahun, C. Catal. Crop yield prediction using machine learning: A systematic literature review. *Comput. Electron. Agric.* **2020**, 177 (July), 105709.
3. D.J. Reddy, M.R. Kumar. Crop yield prediction using machine learning algorithm. *Proc. - 5th Int. Conf. Intell. Comput. Control Syst. ICICCS 2021* **2021**, No. Iciccs, 1466–1470.
4. T. Adugna, W. Xu, J. Fan. Comparison of Random Forest and Support Vector Machine Classifiers for Regional Land Cover Mapping Using Coarse Resolution FY-3C Images. *Remote Sens.* **2022**, 14 (3), 1–22.
5. N. Kussul, M. Lavreniuk, S. Skakun, A. Shelestov. Deep Learning Classification of Land Cover and Crop Types Using Remote Sensing Data. *IEEE Geosci. Remote Sens. Lett.* **2017**, 14 (5), 778–782.
6. M. Drusch, U. Del Bello, S. Carlier, et al. Sentinel-2: ESA’s Optical High-Resolution Mission for GMES Operational Services. *Remote Sens. Environ.* **2012**, 120, 25–36.
7. R. Torres, P. Snoeij, D. Geudtner, et al. GMES Sentinel-1 mission. *Remote Sens. Environ.* **2012**, 120, 9–24.
8. L. Viskovic, I.N. Kosovic, T. Mastelic. Crop classification using multi-spectral and multitemporal satellite imagery with machine learning. *2019 27th Int. Conf. Software, Telecommun. Comput. Networks, SoftCOM 2019* **2019**, 1–5.
9. A. Verma, A. Kumar, K. Lal. Kharif crop characterization using combination of SAR and MSI Optical Sentinel Satellite datasets. *J. Earth Syst. Sci.* **2019**, 128 (8).
10. S. Feng, J. Zhao, T. Liu, et al. Crop Type Identification and Mapping Using Machine Learning Algorithms and Sentinel-2 Time Series Data. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2019**, 12 (9), 3295–3306.
11. B.D. Wardlow, S.L. Egbert, J.H. Kastens. Analysis of time-series MODIS 250 m vegetation index data for crop classification in the U.S. Central Great Plains. *Remote Sens. Environ.* **2007**, 108 (3), 290–310.
12. B. Al-Awar, M.M. Awad, L. Jarlan, D. Courault. Evaluation of Nonparametric Machine-Learning Algorithms for an Optimal Crop Classification Using Big Data Reduction Strategy. *Remote Sens. Earth Syst. Sci.* **2022**, 5 (3), 141–153.
13. J. Yao, J. Wu, C. Xiao, Z. Zhang, J. Li. The Classification Method Study of Crops Remote Sensing with Deep Learning, Machine Learning, and Google

Earth Engine. *Remote Sens.* **2022**, 14 (12), 1–22.

14. E. Maggiori, Y. Tarabalka, G. Charpiat, P. Alliez. Convolutional Neural Networks for Large-Scale Remote-Sensing Image Classification. *IEEE Trans. Geosci. Remote Sens.* **2017**, 55 (2), 645–657.
15. R. Khatami, G. Mountrakis, S. V. Stehman. A meta-analysis of remote sensing research on supervised pixel-based land-cover image classification processes: General guidelines for practitioners and future research. *Remote Sens. Environ.* **2016**, 177, 89–100.
16. X. Huang, L. Zhang. An SVM ensemble approach combining spectral, structural, and semantic features for the classification of high-resolution remotely sensed imagery. *IEEE Trans. Geosci. Remote Sens.* **2013**, 51 (1), 257–272.
17. R. Saini, S.K. Ghosh. CROP CLASSIFICATION ON SINGLE DATE SENTINEL-2 IMAGERY USING RANDOM FOREST AND SUPPOR VECTOR MACHINE. *Int. Arch. Photogramm. Remote Sens. Spat. Inf. Sci.* **2018**, XLII–5 (November), 20–23.
18. P.O. Gislason, J.A. Benediktsson, J.R. Sveinsson. Random forests for land cover classification. *Pattern Recognit. Lett.* **2006**, 27 (4), 294–300.
19. X. Fan, C. Yan, J. Fan, N. Wang. Improved U-Net Remote Sensing Classification Algorithm Fusing Attention and Multiscale Features. *Remote Sens.* **2022**, 14 (15), 1–24.
20. S. Wei, H. Zhang, C. Wang, Y. Wang, L. Xu. Multi-temporal SAR data large-scale crop mapping based on U-net model. *Remote Sens.* **2019**, 11 (1).
21. Y. Wang, J. Su, X. Zhai, F. Meng, C. Liu. Snow Coverage Mapping by Learning from Sentinel-2 Satellite Multispectral Images via Machine Learning Algorithms. *Remote Sens.* **2022**, 14 (3), 1–19.
22. P. Nevavuori, N. Narra, T. Lipping. Crop yield prediction with deep convolutional neural networks. *Comput. Electron. Agric.* **2019**, 163 (April), 104859.
23. C. Fan, R. Lu. UAV image crop classification based on deep learning with spatial and spectral features. *IOP Conf. Ser. Earth Environ. Sci.* **2021**, 783 (1).
24. B. Liu, Y. Shi, Y. Duan, W. Wu. UAV-based crops classification with joint features from orthoimage and DSM data. *Int. Arch. Photogramm. Remote Sens. Spat. Inf. Sci. - ISPRS Arch.* **2018**, 42 (3), 1023–1028.
25. Z. Wang, Z. Zhao, C. Yin. Fine Crop Classification Based on UAV Hyperspectral Images and Random Forest. *ISPRS Int. J. Geo-Information* **2022**, 11 (4).
26. J. Wei, Y. Cui, W. Luo, Y. Luo. Mapping Paddy Rice Distribution and Cropping Intensity in China from 2014 to 2019 with Landsat Images, Effective Flood Signals, and Google Earth Engine. *Remote Sens.* **2022**, 14 (3).
27. A. Stoian, V. Poulain, J. Inglada, V. Poughon, D. Derksen. Land cover maps production with high resolution satellite image time series and convolutional neural networks: Adaptations and limits for operational systems. *Remote Sens.* **2019**, 11 (17), 1–26.
28. A. Soni, R. Koner, V.G.K. Villuri. M-UNet: Modified U-Net Segmentation Framework with Satellite Imagery BT - Proceedings of the Global AI Congress 2019; Mandal, J. K., Mukhopadhyay, S., Eds.; Springer Singapore, Singapore, **2020**; pp 47–59.
29. X. Zhang, Y. Zhou, J. Luo. Deep learning for processing and analysis of remote sensing big data: a technical review. *Big Earth Data* **2022**, 6 (4), 527–560.
30. A. Vali, S. Comai, M. Matteucci. Deep learning for land use and land cover classification based on hyperspectral and multispectral earth observation data: A review. *Remote Sens.* **2020**, 12 (15).
31. P. Ulmas, I. Liiv. Segmentation of Satellite Imagery using U-Net Models for Land Cover Classification. **2020**, 1–11.
32. S. Wang, W. Chen, S.M. Xie, G. Azzari, D.B. Lobell. Weakly supervised deep learning for segmentation of remote sensing imagery. *Remote Sens.* **2020**, 12 (2), 1–25.
33. O. Ronneberger, P. Fischer, T. Brox. U-Net: Convolutional Networks for Biomedical Image Segmentation BT - Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015; Navab, N., Hornegger, J., Wells, W. M., Frangi, A. F., Eds.; Springer International Publishing, Cham, **2015**; pp 234–241.
34. F.I. Diakogiannis, F. Waldner, P. Caccetta, C. Wu. ResUNet-a: A deep learning framework for semantic segmentation of remotely sensed data. *ISPRS J. Photogramm. Remote Sens.* **2020**, 162 (March 2019), 94–114.
35. Wkentaro. labelme. 2021.
36. S. Cai, Y. Tian, H. Lui, et al. Dense-unet: A novel multiphoton in vivo cellular image segmentation model based on a convolutional neural network. *Quant. Imaging Med. Surg.* **2020**, 10 (6), 1275–1285.
37. C. Balakrishna, S. Dadashzadeh, S. Soltaninejad. Automatic detection of lumen and media in the IVUS images using U-Net with VGG16 Encoder. **2018**, 1–10.
38. Z. Zhou, M.M.R. Siddiquee, N. Tajbakhsh, J. Liang. UNet++: A Nested U-Net Architecture; Springer International Publishing, **2018**; Vol. 11045 LNCS.
39. W. Zhu, N. Zeng, N. Wang. Sensitivity, specificity, accuracy, associated confidence interval and ROC analysis with practical SAS® implementations. *Northeast SAS Users Gr. 2010 Heal. Care Life Sci.* **2010**, 1–9.
40. C.M. Florkowski. Sensitivity, specificity, receiver-operating characteristic (ROC) curves and likelihood ratios: communicating the performance of diagnostic tests. *Clin. Biochem. Rev.* **2008**, 29 Suppl 1 (August), S83-7.