IRS LISS – III multispectral data semantic segmentation using deep learning

Dr. Nirav Desai
D-UIAS AND D-SIM&C, Valsad, INDIA
Email: niravdesai.research@gmail.com

Dr. Parag Shukla
School of cyber security and digital forensics, NFSU, Gandhinagar, INDIA

Dr. Akruti Naik
D-UIAS AND D-SIM&C, Valsad, INDIA

ABSTRACT - Identification of land use/land cover and measurement of area under cultivation for various classes is a very significant job. Indeed, the methods and tools used for land use and land cover analysis often require significant human effort, consume substantial time, and can be very expensive. Remote Sensing (RS) plays a pivotal in addressing these challenges efficiently and effectively. Remote sensing provides significant and distinguished data. The current investigation performs semantic segmentation of multispectral remote sensing images using a fully convolutional network (FCN). A multispectral image captures image data within specific wavelength ranges across the electromagnetic spectrum (EM). It has more than 100 nm resolution and less the 10 bands. Semantic Segmentation aims at a pixel-level classification of RS images where every individual pixel is categorized into a distinct class. A total of 4 classes were successfully identified in the present study. Identified classes are water bodies, uncultivated land, residential area and vegetation. Deep learning algorithms U-Net and DeepLabv3+ were used to perform classification on 2 datasets of different sizes and seasons (Dataset – 1:1470 images, Dataset – 2:13500 images). U-Net outperforms DeeplabV3+ in terms of performance. U-net achieved an accuracy of 81 % for dataset - 1, and 84% for dataset - 2 respectively whereas Deeplabv3+ achieved an accuracy of 36% for dataset -1, and 29% for dataset - 2.

KEYWORDS: remote sensing, land use land cover classification, deep learning, U-Net, Deeplabv3+

1. INTRODUCTION

Lately, Remote Sensing (RS) imagery has emerged as a significant source of information for investigative the spatial and time-based changes in ecological situations. Remote sensing (RS) involves the skill of retrieving and comprehending information from distant sources through sensor technology, without direct communication with the observed object [22]. The classification of land use and land cover aims to categorize satellite images into specific classes, based on the distribution of identified land use and land cover categories. Effective land use and land cover mapping are crucial for planning and management strategies. Artificial intelligence techniques, such as neural networks, K-means clustering, Random Forests, and Support Vector Machines, are actively used for classifying remote sensing imagery, utilizing pattern recognition and computer vision methodologies. Land use and land cover classification using remote sensing images has been applied in numerous studies, including surveys focused on environmental monitoring and change

detection. The introduction of digital multispectral imagery in remote sensing has revolutionized the mapping and monitoring of Earth's natural resources, as well as land use and land cover analysis [10]. These images can also be used for urban areas analysis [3]. Remote sensing is recognized as a very cost-effective tool for mapping or classification of land use and land cover [23]. The classification of land use land cover can be achieved through supervised or unsupervised classification methods. The aim behind the classification of multispectral images is used to extract data from remotely sensed data and the recognition of homogeneous regions in an image. Remote sensing system's spatial resolution is typically limited which can sometimes make it challenging to differentiate amongst numerous classes on the Earth's surface [24]. [8] projected deep learning and the training challenges associated with deep neural networks can potentially be mitigated through a one-by-one layer initialization approach. And this approach was effectively applied to the domains of video and image processing, both of which are crucial areas within the realm of data analysis [3]. Deep learning algorithms have presented a novel approach to remote sensing image interpretation, leading to a substantial body of research in the realm of remote sensing image classification. Semantic Segmentation aims at a pixel-level classification of RS images where every individual pixel is categorized into a distinct class. A multitude of machine learning and deep learning models are offered to tackle segmentation tasks effectively.

The study was conducted in the South Gujarat region of Gujarat state, India, focusing on the classification of four distinct classes: Water Bodies, Vegetation, Uncultivated Land, and Residential Areas. Semantic segmentation, which involves pixel-level classification by assigning each pixel to its respective class, was employed using deep neural networks.

The study performed the semantic segmentation on 2 dissimilar datasets (Dataset -1 of 1470 images and Dataset -3 of 13500 images) of LISS -III remote sensing multi spectral data images (South Gujarat Region, INDIA). A total of 4 classes were successfully identified in the present study. Identified classes are water bodies, uncultivated land, residential area and vegetation. U-Net and DeepLabv3+ were utilized for the classification of remote sensing images. U-Net outperforms DeeplabV3+ in terms of performance. U-net achieved an accuracy of 81 % for dataset -1, and 84% for dataset 02 respectively, whereas Deeplabv3+ achieved an accuracy of 36% for dataset -1 and 29% accuracy for dataset – 2.

2. LITERATURE REVIEW

Deep learning (DL) is a subset of machine learning techniques categorised by its utilization of multiple layers in organized architectures, enabling unsupervised learning and pattern classification [17]. Huang [7] showed classification using both pixel and object-based methods and achieved 87% accuracy when investigating satellite images. Minaee [15] presented a deep learning-based DFS algorithm and directed a comparison with YOLOv2. The dataset utilized in the experiment was sourced from Google Earth and involved the detection of a total of 6 objects, yielding an achievement of 85% accuracy. [23] used Maximum likelihood for classification in the experiment. Images acquired from SPOT, ASTER, and Landsat TM sensors were used and obtained Overall Accuracy (OOA) for the respective years was 85.33% in 1989, 86.67% in 2000, 88.33% in 2010, and 86.38% in 2016. [6] focused on Saliency Dual Attention Residual for the purpose of achieving high-performance results. Lateef [12], the proposed classification technique involves the combination of a neural network with random forest (RF) for land use and land cover classification. Moreira [18], numerous strategies were projected to improve the performance of convolutional neural networks (CNNs) for the segmentation of aerial images. Semantic segmentation is the task of assigning each pixel in an image to specific predefined classes. It aims to identify the substances present in the image and determine their locations [19]. Deep Learning (DL) and Convolutional Neural Networks (CNN) have experienced a significant revolution in the field of image classification over an extended period, emerging as central methods for this task [10]. Li [14] provides a structured and comprehensive assessment of different approaches in the application of Deep Learning (DL) methods. Ding [4] emphasized the importance of diverse training strategies and techniques during the training process. [25] provides a condensed overview of various

techniques, while [11] and [16] offer in-depth coverage of new advances in the field.

3. MATERIALS AND METHOD

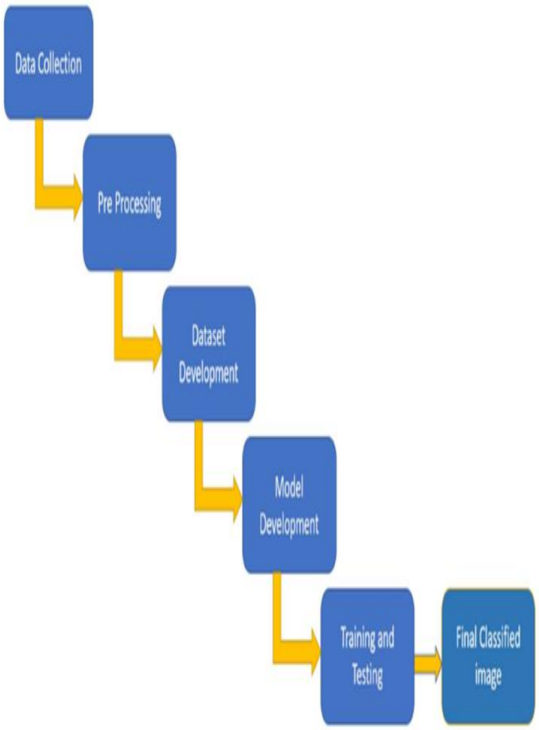

Figure 1: Research Methodology

Figure 1 illustrates the research methodology for conducting land use land cover classification using an IRS LISS – III multispectral image. The methodology includes several key steps: data collection, preprocessing of the imagery, development of the dataset, creation and training of the classification model, testing its accuracy, and finally producing the classified image depicting different land use and land cover types.

3.1 DATA COLLECTION

Indian Remote Sensing (IRS) LISS- III Multispectral remote sensing images were used to experiment. Multispectral images typically consist of fewer than or equal to 10 bands and offer a resolution finer than 100 nanometres (nm). The IRS LISS-III multispectral remote sensing image is composed of four separate bands stored in distinct .tiff files. These bands correspond to Band 2 (Blue), Band 3 (Green), Band 4 (Red), and Band 5 (Near Infrared). The spatial resolution is 30 meters and Quadrats of 30m * 30m size were placed across the study area. The LISS-III data was sourced from https://bhuvan-app3.nrsc.gov.in and subsequently processed using the ENVI 4.7 image analysis software. A widespread ground study was accomplished to gather environmental landscapes and circulation forms of dissimilar land use and land cover. Ground Control Points (GCPs) were collected by recording the Longitude and Latitude coordinates of specific locations for their respective classes. This data collection process was facilitated with the use of the GPS device Garmin eTrex 30. The allocation of Ground Control Points (GCPs) for each specific category was determined based on the spatial distribution of identified land use and land cover classes within the study area.

3.2 PRE-PROCESSING

To identify land use and land cover classes in the study, the process involved creating a False Colour Composite (FCC) image by stacking the IRS LISS-III multispectral images. This FCC was formed by grouping Band-4, Band-3, and Band-2. Subsequently, ground truth masks were generated for each image after the construction of the FCCs, and these masks were then used for training the deep learning model. This approach allowed for the development of a supervised learning model that could classify and map different land use and land cover categories based on the provided ground truth information.

3.3 DATASET

The study was performed on a total of 2 datasets which contain the FCC image of different seasons and Ground Truth Masks. The process involved generating False-Color Composites (FCCs) by stacking up .TIFF image files, specifically Band-4, Band-3, and Band-2, to create the FCC images. Ground truth masks were then created for each class using the Maximum Likelihood (ML) classifier within the region of interest. Both the FCC images and their corresponding masks were resized to 1024 x 1024 pixels.

Figure 2: FCC Image

There were two datasets:

Dataset - 1 consisted of 1470 images, with 1,255 images allocated for training the deep learning model, while the remaining 215 images were set aside for validation and evaluation.

Dataset - 2 had a larger size of 13,500 images, with 11,250 images designated for model training, and the remaining 2,250 images reserved for validation and evaluation.

These datasets were used to train and assess the performance of the deep learning model, likely for tasks related to land use and land cover classification or image segmentation.

3.4 METHODOLOGY

The Maximum Likelihood (ML) classifier was applied to the LISS-III multispectral image data. This classifier assigns each pixel to the class that has the maximum likelihood of being a match. In other words, for each pixel, the ML classifier calculates the probability of it belonging to each class and assigns it to the class with the highest probability. This statistical approach is based on the likelihood of observed pixel values given the statistical characteristics of each class and is commonly used for image classification tasks. The mean vector and covariance matrix are fundamental components of the Maximum Likelihood (ML) classifier that can be enhanced through training data [22]. round truth masks were generated by applying the Maximum Likelihood (ML) algorithm to the region of interest for each class. This process involved using the ML classifier to classify the pixels within the defined regions of interest into their respective land use or land cover classes, thereby creating accurate reference masks for training and evaluation purposes. Figure 3 represents the ground truth mask Figure 4 illustrates the fundamental principles of Maximum Likelihood [5]. Following the creation of False Colour Composite (FCC) images, ground truth masks were generated and subsequently employed for training the model. These masks served as essential reference data for teaching the model to accurately classify land use and land cover classes.

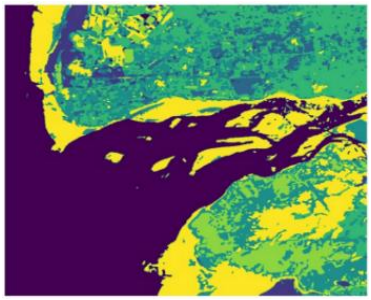

Figure 3: Ground Truth Mask

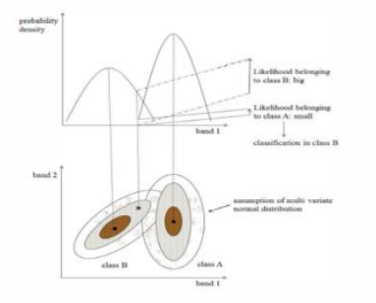

Figure 4: Basic concept of ML

Fully Convolutional Networks (FCNs) have found efficient applications in various fields, including image segmentation [1,7], as well as in medicinal image analysis [5,13]. FCN (Fully Convolutional Network) differs from conventional neural networks in that it processes information more holistically, considering the entire input image within its

convolutional layers rather than being limited to localized receptive fields [9]. This characteristic makes FCN well-suited for tasks like pixel-based classification [20], where it can efficiently capture spatial relationships and contextual information within images to make accurate predictions. DeepLabv3+ is a deep learning model for semantic image segmentation where the goal is to assign semantic labels to every pixel in the input image. The DeepLabv3+ model has an encoding phase and a decoding phase. The encoding phase extracts the important data from the image using a convolutional neural network (CNN) while the decoding phase rebuilds the output of suitable dimensions based on the data obtained from the encoder phase.

Semantic Segmentation is defined as a pixel-level classification of images where a class is allotted to each pixel of the image. In the present study, there are 4 classes - Water Bodies, Vegetation, Uncultivated Land, and Residential areas. A deep neural network was used to handle this task. The architecture described is a U-Net-based Fully Convolutional Network (FCN) with skip connections, designed for semantic segmentation tasks.

FCNs utilize an encoder for feature extraction and a decoder for reconstruction. In this experiment, a fully-convolutional network with skip connections was trained to process image inputs of size 256 x 256 x 3 and generates output matrices of shape 256 x 256 x 4, representing a one-hot encoded version of the segmentation mask. The U-Net architecture includes both an encoder and a decoder. The encoder consists of 5 blocks, each composed of 2 layers—convolution, batch normalization, and ReLU activation—stacked together, followed by max-pooling in all but the last block. The output of the encoder is then fed into the decoder, which comprises 4 blocks. Each decoder block begins with upsampling of the input, followed by a 1 x 1 convolution operation. Skip connections concatenate the output of the corresponding encoder block with the output of the upsampling and convolution operation in the decoder. This concatenated tensor is further processed through two convolutional layers similar to those in the encoder blocks. Finally, the output of the decoder undergoes a 1 x 1 convolution operation with 4 filters, corresponding to the number of classes (4).

Here's a breakdown of the key components and operations:

Input Size: The model takes input images of two different sizes: 256 x 256 x 3 and 128 x 128 x 3.

Encoder Part: The encoder consists of 5 blocks, each containing two convolutional layers followed by batch normalization and ReLU activation. Max-pooling is applied after each block, except for the last one.

Decoder Part: The decoder consists of 4 blocks. Each block in the decoder begins with an upsampling operation followed by a 1 x 1 convolution layer. A skip connection is established by concatenating the output of the corresponding encoder block with the output of the upsampling and convolution operation. The concatenated tensor is then processed through two more convolution layers, similar to the corresponding encoder block.

Final Layer: The output of the decoder part is passed through a 1 x 1 convolution layer with the number of filters equal to the number of classes, which is 4 in this case.

Figure 5 provides a visual representation of the U-Net architecture, where arrows depict the flow of data, black containers represent feature maps, and gray containers represent cropped feature maps from the contracting path. This architecture is particularly effective for image segmentation tasks, where it can capture both local details and global context through the use of skip connections and skip concatenations

$$E = \sum w(x)\log\left(P_{k(x)}(x)\right) \quad (1)$$

Where pk is the pixel-wise SoftMax function applied over the final feature map.

$$P_k(x) = \frac{e^{ak(x)}}{\sum_{k!=}^{k} eak(x)}, \quad (2)$$

And ak (x) denotes the activation in channel k.

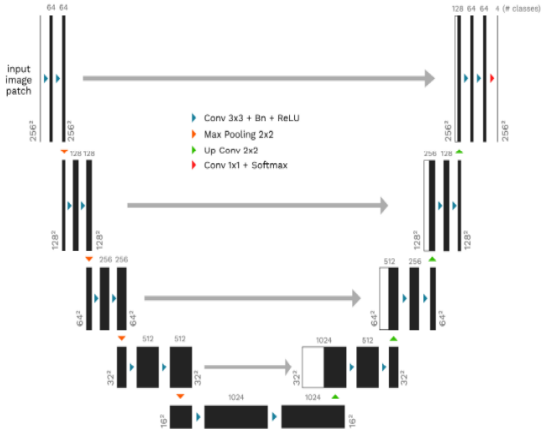

Figure 5: U-Net architecture

DeepLabv3+ is a cutting-edge semantic segmentation model characterized by its encoder-decoder architecture. The encoder incorporates a pre-trained CNN model to generate encoded feature maps from the input image, while the decoder reconstructs the output based on the crucial information gleaned by the encoder through upsampling techniques. The architecture of DeepLabv3+ is depicted in Figure 6 [2]. DeepLabv3+ builds upon DeepLabv3 through the incorporation of an encoder-decoder architecture. In this enhanced model: The encoder component excels at handling multiscale contextual information. It achieves this by employing dilated convolutions across various scales. This approach enables the network to capture a broader context without introducing a significant increase in parameters or computational complexity. The decoder module plays a critical role in refining the segmentation outcomes, particularly in cases involving long object boundaries. It complements the encoder's work by further enhancing the quality of the segmentation output. This architecture leverages dilated convolutions effectively to maintain a consistent stride while significantly expanding the field-of-view. By doing so, it can provide more comprehensive context for pixel-wise semantic segmentation, ultimately enhancing the accuracy of the results.

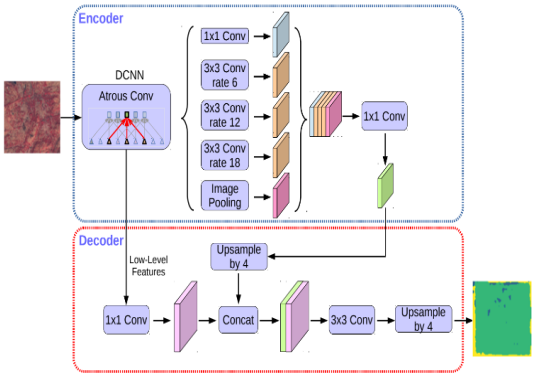

Figure 6 Deeplabv3+ architecture

3.5 ALGORITHM STEPS

The workflow for the image segmentation or classification process can be summarized in the following steps:

Pre-processing: Initially, pre-processing steps are applied to the remotely sensed images. This includes image alteration, registration, and masking. In this case, the images consist of bands 2 to band 4.

Class Selection: Four specific classes are chosen for training and testing purposes. These classes include Water Bodies, Vegetation, Uncultivated Land, and Residential Areas. ENVI image processing software, specifically the ROI (Region of Interest) Tool, is employed to select and define these classes within the images.

Model Building and Training: Four selected classes are used to build and train the segmentation or classification model.. The model's architecture and parameters are configured during this phase and are preserved for future image segmentation or classification tasks after training.

Model Application: The model that was trained in the previous step is then applied to perform classification on the target images.

Outcome Verification: The algorithm checks whether all the classifications have been completed or not. If the classification process has been finished for all the desired areas or regions, the classification results are displayed. This marks the end of the algorithm.

This workflow outlines the key steps involved in the image segmentation or classification process, from initial data preparation to the final classification outcomes.

3.6 TRAINING CONFIGURATION

During the data ingestion process, the input images were standardized by clipping their pixel values to the range [0.0, 255.0]. For the masks, they were one-hot encoded to match the total number of classes. Random augmentations were applied to both the images and masks before feeding them into the model for training. These augmentations helped expand the dataset and increased the model's robustness to handle different orientations and variations beyond what was present in the training data. Properly applied augmentations also served as a preventive measure against over fitting. To support this data ingestion pipeline, a custom image data generator was developed, ensuring that the data met the necessary requirements for training.
Table 1 outlines the hyper parameters and other configurations used during the model training, providing essential details for reproducing or understanding the training process.

Hyper parameter/Configuration	Value
Image normalization	Clip images
Mask encoding	One-hot encoding (4 classes)
Data augmentation	Random augmentations
Custom data generator	Implemented
Train Batch Size	16
Validation Batch Size	16
Input Image Shape	(256,256,3)
Number of classes	4
Epochs	50
Loss	Categorical Focal Loss*
Optimizer	Adam
Metrics	Dice Coefficient*
Class Weights	[1.69941,0.53043,1.23977, 1.38949]

Table 1: Hyper parameters and Training Configurations

4. RESULT

The experiment was conducted on a Windows 11 operating system using Envi 4.7 to process remote sensing information, including tasks such as selecting training samples and generating masks. The classification algorithm, implemented in U-Net and Tiramisu architectures, was developed using Python and OpenCV. Figure 5 illustrates the land cover classification results for the South Gujarat region, India, with Water Bodies depicted in black, Vegetation in light green, Uncultivated Land in light blue, and Residential Areas in yellow.

The classification algorithm, implemented using U-Net and DeepLabv3+, was coded using Python and the OpenCV library. The experimental phase aimed to fine-tune the parameters for U-Net and DeepLabv3+ using the RGB bands of the IRS LISS-III multispectral remote sensing image dataset. The objective was to identify the parameter configurations that yielded the best performance. The chosen parameters, which demonstrated excellent performance during experimentation, were selected to develop the final model. Additionally, data augmentation techniques were applied to the dataset to enhance the model's ability to generalize and make accurate predictions on diverse data samples. Overall, this approach involved iteratively tuning and refining the model parameters until achieving optimal results for the specific task of classifying remote sensing images with the RGB bands from the IRS LISS-III dataset.

Model	Optimizer	EPOCH Trained	Dataset	Accuracy %
U-net	Adam	50	Dataset - 1	81
U-net	Adam	50	Dataset - 2	84
Deeplabv3+	Adam	50	Dataset - 1	36
Deeplabv3+	Adam	50	Dataset - 2	29

Table 2: Experiment result

Table 2 shows the experiment results and accuracy with different epochs. And from the outcomes, it was detected that U-net gives better results in classifying

land use land cover classes. Figure 6 shows the results predicted by U-Net and Deeplabv3+ models. Figures 7(a) and 7(b) displays the land use land cover classification by the U-Net model on dataset -1 and dataset -2 respectively. Figures 7(c) and 7(d) show the classification of land use land cover by Deeplabv3+ on dataset -1 and dataset -2. The resultant images shown Water Bodies in black color, Vegetation in light green color, Uncultivated Land in light blue color, and Residential Areas in yellow color.

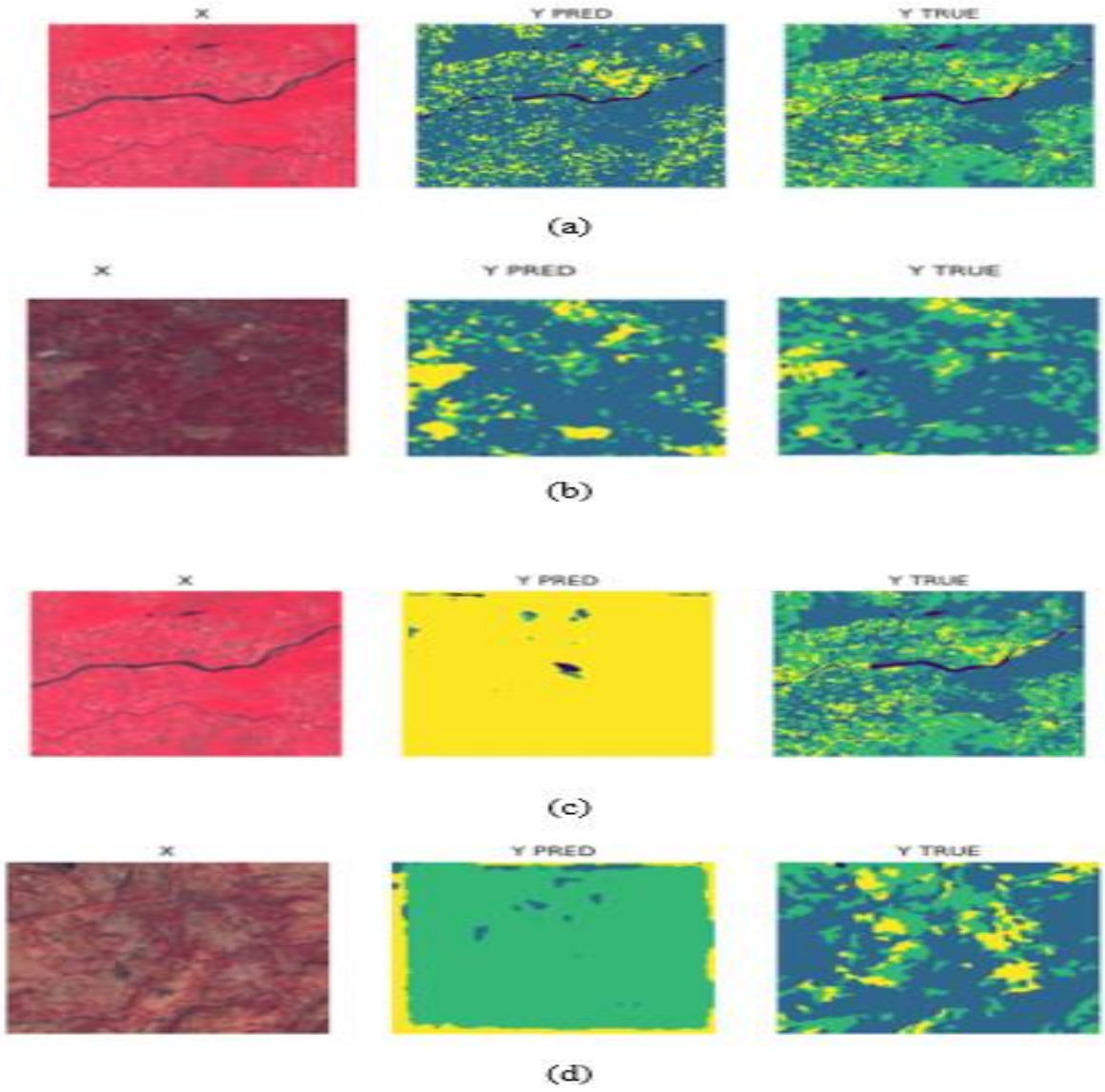

Figure 7: Classified images

Figure 8 illustrates the quantification results for each respective land use and land cover class following the completion of the classification process. The numerical values assigned to each class are 0: Water,

1: Vegetation, 2: Uncultivated Land ,3: Residential Area. These values represent the classification labels assigned to different regions or pixels within the remote sensing images, allowing for a detailed assessment of the land use and land cover distribution in the study area.

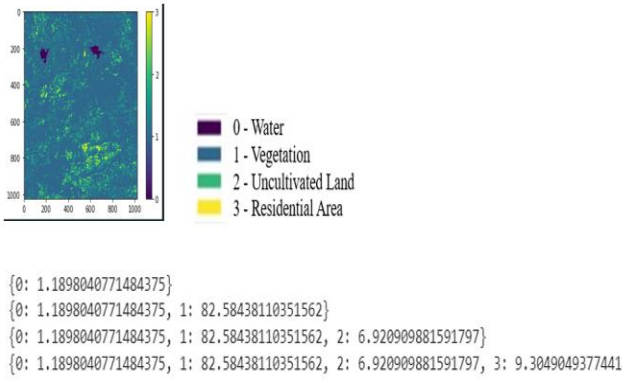

Figure 8: Quantifications for each class

Table 3 represents class values, generated by U-Net classification mechanism in which vegetation cover maximum area and water covers least area by percentage cover.

Class Value	Class Name	Percentage cover in image
0	Water	1
1	Vegetation	44
2	Uncultivated Land	23
3	Residential area	32

Table 3: Quantification of resultant image

5. CONCLUSION

The images predicted by the U-Net closely resemble the ground truth masks. This research was conducted in the South Gujarat region of India, encompassing data captured during different months and seasons (October, January, and May). The proposed model has demonstrated impressive accuracy in Land Use Land Cover (LULC) classification using deep learning techniques, highlighting the increasing importance of deep learning in this domain. Compared to traditional visual interpretation methods, deep learning offers a cost-effective and efficient solution for LULC classification. In this study, we introduced a novel LULC classification model utilizing the U-Net and Deeplabv3+ algorithms, trained and evaluated on the LISS-III multispectral satellite imagery dataset. Our experiments successfully identified four primary land use land cover classes: Water Bodies, Vegetation, Uncultivated Land, and Residential Areas, achieving high accuracy.

The combination of Maximum Likelihood for generating ground truth masks and U-Net for classification has proven to be effective, delivering superior results for your specific land use and land cover classification task. This approach likely leveraged the strength of Maximum Likelihood for accurate ground truth labelling and U-Net's capabilities in handling pixel-wise classification tasks, resulting in improved overall performance and accuracy in classifying the remote sensing images. The result shows that the U-net performed better than Deeplabv3+ on various kinds of datasets. Furthermore, present study proved that deep learning is better for land use land cover classification. Deep learning is providing a more cost-effective result than the visual interpretation or other machine learning techniques of today. This model was trained and tested on two datasets comprised of IRS LISS-III multispectral images. The experimental results demonstrated that the model is capable of accurately detecting a total of four land use and land cover classes with high accuracy. These findings confirm that the FCN classifier holds significant potential for the precise identification of land use and land cover classes using remote sensing techniques. This achievement underscores the effectiveness of deep learning approaches in remote sensing applications. The results indicate that the U-Net classifier outperforms Deeplabv3+, highlighting its significant potential for accurately detecting land use land cover classes. Specifically, the U-Net model achieved impressive accuracy, while Deeplabv3+ achieved moderate accuracy. The deep learning algorithms U-

Net and DeepLabv3+ were employed for classification on two datasets differing in size and seasonality (Dataset 1: 1470 images, Dataset 2: 13500 images). U-Net consistently outperformed DeepLabv3+ in terms of accuracy. Specifically, U-Net achieved 81% accuracy for Dataset 1 and 84% accuracy for Dataset 2, while DeepLabv3+ achieved 36% accuracy for Dataset 1 and 29% accuracy for Dataset 2.

REFERENCES

1. Bi, L., Kim, J., Ahn, E., Kumar, A., Feng, D., & Fulham, M. (2019). Step-wise integration of deep class-specific learning for dermoscopic image segmentation. *Pattern recognition*, *85*, 78-89.
2. Chen, L. C., Papandreou, G., Kokkinos, I., Murphy, K., & Yuille, A. L. (2017). Deeplab: Semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected crfs. *IEEE transactions on pattern analysis and machine intelligence*, *40*(4), 834-848.
3. Dian-lai, W., Ai-xia, S., & Wen-ping, L. (2020, November). Application of deep neural networks in classification of medium resolution remote sensing image. In *Journal of Physics: Conference Series* (Vol. 1682, No. 1, p. 012014). IOP Publishing.
4. Ding, Y., Zhu, Y., Feng, J., Zhang, P., & Cheng, Z. (2020). Interpretable spatio-temporal attention LSTM model for flood forecasting. *Neurocomputing*, *403*, 348-359.
5. Fan, J., Cao, X., Yap, P. T., & Shen, D. (2019). BIRNet: Brain image registration using dual-supervised fully convolutional networks. *Medical image analysis*, *54*, 193-206.
6. Guo, D., Xia, Y., & Luo, X. (2020). Scene classification of remote sensing images based on saliency dual attention residual network. *IEEE Access*, *8*, 6344-6357.
7. Huang, Y., Zhou, F., & Gilles, J. (2019). Empirical curvelet based fully convolutional network for supervised texture image segmentation. *Neurocomputing*, *349*, 31-43.
8. Hinton, G. E., Osindero, S., & Teh, Y. W. (2006). A fast learning algorithm for deep belief nets. *Neural computation*, *18*(7), 1527-1554.
9. Hoeser, T., Bachofer, F., & Kuenzer, C. (2020). Object detection and image segmentation with deep learning on Earth observation data: A review—Part II: Applications. *Remote Sensing*, *12*(18), 3053.
10. Karoui, M. S., Deville, Y., Hosseini, S., Ouamri, A., & Ducrot, D. (2009, August). Improvement of remote sensing multispectral image classification by using independent component analysis. In *2009 First Workshop on Hyperspectral Image and Signal Processing: Evolution in Remote Sensing* (pp. 1-4). IEEE.
11. Khedam, R., Outemzabet, N., Tazaoui, Y., & Belhadj-Aissa, A. (2006, April). Unsupervised multispectral image classification using artificial ants. In *2006 2nd International Conference on Information & Communication Technologies* (Vol. 1, pp. 349-354). IEEE.
12. Lateef, F., & Ruichek, Y. (2019). Survey on semantic segmentation using deep learning techniques. *Neurocomputing*, *338*, 321-348.
13. Li, C., Wang, X., Liu, W., Latecki, L. J., Wang, B., & Huang, J. (2019). Weakly supervised mitosis detection in breast histopathology images using concentric loss. *Medical image analysis*, *53*, 165-178.
14. Li, L., Han, L., Ding, M., Cao, H., & Hu, H. (2021). A deep learning semantic template matching framework for remote sensing image registration. *ISPRS Journal of Photogrammetry and Remote Sensing*, *181*, 205-217.
15. Minaee, S., Boykov, Y. Y., Porikli, F., Plaza, A. J., Kehtarnavaz, N., & Terzopoulos, D. (2021). Image segmentation using deep learning: A survey. *IEEE transactions on pattern analysis and machine intelligence*.
16. Mishra, C., & Gupta, D. L. (2017). Deep machine learning and neural networks: An overview. *IAES International Journal of Artificial Intelligence*, *6*(2), 66.
17. Mohanty, S. P., Czakon, J., Kaczmarek, K. A., Pyskir, A., Tarasiewicz, P., Kunwar, S.,

... & Schilling, M. (2020). Deep learning for understanding satellite imagery: An experimental survey. *Frontiers in Artificial Intelligence*, *3*, 534696.

18. Moreira, R. C. (2008). *Estudo espectral de alvos urbanos com imagens do sensor HSS (Hyperspectral Scanner System)* (Doctoral dissertation, PhD Thesis, National Institute for Space Research (INPE)).
19. Pires de Lima, R., & Marfurt, K. (2019). Convolutional neural network for remote-sensing scene classification: Transfer learning analysis. *Remote Sensing*, *12*(1), 86.
20. Ptucha, R., Such, F. P., Pillai, S., Brockler, F., Singh, V., & Hutkowski, P. (2019). Intelligent character recognition using fully convolutional neural networks. *Pattern recognition*, *88*, 604-613.
21. Ronneberger, O., Fischer, P., & Brox, T. (2015, October). U-net: Convolutional networks for biomedical image segmentation. In *International Conference on Medical image computing and computer-assisted intervention* (pp. 234-241). Springer, Cham.
22. Sarah C. Goslee" Analyzing Remote Sensing Data in R: The Landsat Package", Journal of Statistical Software, July 20i I, Volume 43, Issue http://www.jstatsoft.orgl.
23. Schowengerdt, R. A. (2006). *Remote sensing: models and methods for image processing*. Elsevier.
24. Xu, X., Chen, Y., Zhang, J., Chen, Y., Anandhan, P., & Manickam, A. (2021). A novel approach for scene classification from remote sensing images using deep learning methods. *European Journal of Remote Sensing*, *54*(sup2), 383-395.
25. Yang, D., Xie, Z., & Du, X. (2020, September). Analysis of Land Cover Change in Shenyang Based on Remote Sensing Image Supervised Classification Technology. In *Journal of Physics: Conference Series* (Vol. 1631, No. 1, p. 012140). IOP Publishing.