ISSN No:-2456-2165

Environmental Exploration and Monitoring of Vegetation Cover using Deep Convolutional Neural Network in Gombe State

Okere Chidiebere Emmanuel, Abdulrauf Abdulrasheed
Department of Computer Science
Federal Polytechnic Kaltungo
Gombe state, Nigeria

Mustapha Abdulrahman Lawal
Department Of Mathematical Sciences
Abubakar Tafawa Balewa University
Bauchi, Nigeria

Ismail Zahraddeen Yakubu
Department of Computing Technologies
SRM Institute of Science and Technology
Kattankulathur, Chennai, India, 603203

Abstract:- In the previous years, human assessments of satellite data were possible because to the relatively modest volume of images accessible; however, this is no longer the case. Additionally, traditional software like ARGIS, EDARS, ILWIS, and other time-consuming and ineffective tools are heavily used by environmental organizations in Nigeria, such as the National Centre for Remote Sensing. With today's large number of data, relevant information extraction from photos thus becomes a challenge. Differentiating between classes with comparable visual qualities is another problem, as seen when attempting to categorize a green pixel as grass, shrubs, or a tree. However, as seen in other computer vision fields, machine learning approaches have shown to be a strong answer in this case. We proposed the introduction of a novel three-dimensional based architecture that is specialized to multispectral pictures and addresses the majority of the challenging features of Deep Learning for Remote Sensing as a solution to the issues raised. This work's major goal is to create a deeper CNN architecture (U-Net) model that is efficient for semantically segmenting remote sensing imagery with additional multi-spectral feature classes. In comparison to traditional remote sensing software, our innovative 3D CNN architecture can analyze the spatial and spectral components simultaneously with true 3D convolutions. our results in better, faster, and more efficient segmentation.

Keywords:- *Computer Vision, Deep Learning, Semantic Segmentation of Satellite Imagery.*

I. INTRODUCTION

The process of classifying each pixel in an image into one of several predetermined classes or categories is known as semantic segmentation of remote sensing imagery. This method is frequently used to interpret and understand satellite or aerial pictures of natural habitats, urban environments, agricultural fields, and landscapes. Semantic segmentation is an essential technique for applications like land use classification, environmental monitoring, disaster assessment, and urban planning since it gives precise information about the distribution of various land cover categories inside an image.

Due to the rapid advancement of remote sensing (RS) technology, high-resolution remote sensing satellites (like IKONOS, SPOT-5, World View, and Quick Bird) produce imagery that is more abundant in information than low-resolution remote sensing imagery, making it easier to extract features and identify ground objects. Today, it is possible to detect a wide variety of artificial things that were previously challenging to identify [1]. Many applications use semantic segmentation of remote sensing imagery, which has been a major area of study for decades. In computer vision (CV) and remote sensing, semantic segmentation has received a lot of attention, relying on shallow features manually generated by experts [2]. Therefore, a framework that works for one task may not work for another depending on the circumstances, resulting in a costly and lengthy rewrite of the entire feature extraction process. These drawbacks prompted academics in the field to hunt for a more reliable and successful strategy [3]. The deep learning research community has reached a state of the art in automating visual labeling tasks.

Researchers have worked hard to apply deep learning techniques' improved performance to the study of remote sensing image processing in light of their success in the field of computer vision [1]. In the last ten years, deep learning techniques have proven to perform significantly better in many common computer vision tasks, such as semantic segmentation and object categorization. Deep learning approaches are more suited to handle complex situations since they automatically produce features that are suitable for specific classification tasks. The enormous success of deep learning techniques in other fields encouraged their adoption and expansion for use in remote sensing problems [1]. Despite decades of work, the literature review [1] reveals the following significant issues that call for research and the creation of fresh approaches: (3) a shortage of training examples, (4) the need for pixel-level accuracy, and (5) the processing of unusual data. This creates a lot of space for more research to address the problems highlighted above.

ISSN No:-2456-2165

Because there were very few photos accessible in the last ten years, human assessments of satellite data were practical. This is no longer the case. Additionally, traditional software like ARGIS, EDARS, ILWIS, and other time-consuming and ineffective tools are heavily used by environmental organizations in Nigeria, such as the National Centre for Remote Sensing. With today's large number of data, relevant information extraction from photos thus becomes a challenge. Differentiating between classes with comparable visual qualities is another problem, as seen when attempting to categorize a green pixel as grass, shrubs, or a tree. However, as seen in other computer vision fields, machine learning approaches have shown to be a strong answer in this case We proposed the introduction of a novel three-dimensional-based architecture that is specialized to multispectral pictures and addresses the majority of the challenging features of Deep Learning for Remote Sensing as a solution to the issues raised. This work's major goal is to create a deeper CNN architecture (U-Net) model that is efficient for semantically segmenting remote sensing imagery with additional multi-spectral feature classes. In comparison to traditional remote sensing software, our innovative 3D CNN architecture can analyze the spatial and spectral components simultaneously with actual 3D convolutions. our results in better, faster, and more efficient segmentation.

II. RELATED WORK

In this study, semantic segmentation and pixel-wise classification are used interchangeably. The term "semantic segmentation" is becoming frequently used in computer vision, and it is also being used more frequently in remote sensing. Modern convolution and segmentation sub-networks are used in end-to-end trained semantic segmentation frameworks for RGB imagery. Due to its superior performance compared to that of conventional learning algorithms, deep learning (DL) has recently emerged as the big data analysis trend that is expanding the fastest. It has been widely and successfully applied to a variety of computer application fields, including speech recognition, natural language processing, sequential data, and image classification [4]. Machine learning techniques are gaining importance as science shifts its focus to data-intensive research. In particular, deep learning has made significant advancements in the sector.

For the mapping of trees, shade, buildings, and roads, Praveena and Singh [5] presented a hybrid clustering technique and feed-forward neural network classifier. The suggested method outperformed every other existing algorithm used as a benchmark. However, the outcomes show that Moving KFCM outperforms the currently used algorithms for classifying shade regions. The performance of a powerful deep neural network was also proposed by [6] and when compared to SIFT, SURF, SAR-SIFT, and PSO-SIFT, the results of the experiments reveal that the use of transfer learning increases accuracy and lowers training costs. But the main drawback of this research is that distinct source photos do not share a common feature representation.

In order to detect water in satellite images for flood assessment, Jony, Woodley [7] uses an ensemble classifier. After evaluating this method against Mediaeval 2017, it was discovered that this method can produce good classification accuracy for both a seen location when bands are used and an unseen location when NDWI is used. The study's biggest flaw, though, was that it produced poorer results on a hidden site.

Furthermore, [8] put forth a brand-new convolutional neural network (CNN) to categorize snow and clouds at the object level. In particular, a novel CNN structure that can learn multiscale semantic aspects of clouds and snow from high-resolution multispectral data is described. The author extends a straightforward linear iterative clustering approach for segmenting high-resolution multispectral pictures and producing super pixels to address the problem of "salt-and-pepper" in pixel-level predictions. Results showed that the new proposed method performs better than the previous methods in terms of accuracy and robustness in separating snow and clouds in high-resolution images. The study, however, fails to apply the suggested convolutional neural network-based techniques to a different objective in the realm of remote sensing, such as urban water extraction and ship detection.

By proposing an end-to-end segmentation model that combines convolution and pooling operations and is capable of learning global relationships between object classes more effectively than conventional classification methods, the study [9] illustrated the utility of FCN architectures for the semantic segmentation of remote sensing MSI. The outcome shown that superior classification performance was achieved for fourteen of the eighteen classes in RIT-9 using an end-to-end semantic segmentation approach. However, the outcome can be enhanced by investigating more intricate ResNet and U-Net models. Additionally, adding more different classes to the synthetic data should help the creation of frameworks that are more discriminative and produce better results.

Remote sensing has previously shown to be a very useful and effective technology for mapping slums. The work in [10] examines how well FCNs can learn from slum mapping in various satellite pictures. Sentinel-2 and TerraSAR-X data are modeled using QuickBird's extremely high-resolution optical satellite images. Although free Sentinel-2 data is freely available, slum mapping is a difficult task due to its considerably poor resolution. On the other hand, TerraSAR-X data is thought to be an effective data source for intra-urban structure study because it has a greater resolution However, transferring the model did not increase the performance of semantic segmentation due to the different image features of SAR compared to optical data, yet we detected extremely high accuracies for mapped slums in the optical data.

To overcome the performance issues with VHR picture semantic segmentation. A Superpixel-enhanced Deep Neural Forest (SDNF) was suggested by [11]. In order to balance the classification capacity and representation learning capability of DCNNs, a fully differentiable forest is implemented to take control of deep convolutional layer representation learning. It is also suggested to use a Super pixel-enhanced Region Module (SRM) to reduce classification noise and improve the boundaries of ground objects. The ISPRS 2D labeling standard is used to gauge SDNF's effectiveness. Results from the experiments show that our technique achieves new state-of-the-art performance.

Similar to this, [3] suggested an FCN-based model to implement pixel-wise classifications for remote sensing images in an end-to-end manner. Additionally, [3] proposed an adaptive threshold approach to alter the threshold of the Jaccard index in each class. The suggested method generates accurate classifications in a comprehensive manner, according to tests on the DSTL dataset. Results indicate that the adaptive threshold approach can improve segmentation accuracy by raising the average Jaccard index score from 0.614 to 0.636. The model's training under weak supervision, which would increase its applicability, is the research's main inadequacy.

A weed/crop segmentation network that was recently proposed in 2020 [12] delivers higher performance for accurately recognizing the weed with arbitrary shape in challenging environmental conditions and offers excellent support for autonomous robots to successfully minimize the density of weeds. By incorporating four extra components and testing the network's performance on the two difficult Stuttgart and Bonn datasets, the deep neural network (DNN)-based segmentation model achieves sustained improvements. The two dataset’s cutting-edge performance demonstrates that each additional component has a significant potential to improve segmentation accuracy. The research does not, however, model the network to learn more correlated spatial information or take advantage of the domain expertise in weed detection.

Due to their excellent spatial resolution and adaptability in picture capture, drones have recently revolutionized the mapping of wetlands. As a result, [13] suggested utilizing image segmentation to map the vegetation in wetlands. To do this, ML and DL algorithms were put to the test on a collection of drone photographs taken of Ireland's Clara Bog, an elevated bog. Overall, the DL approaches' accuracy was around 4% better than the ML techniques. Furthermore, the DL technique does not need any color adjustments or additional textural qualities. However, DL needs a lot of initial labeled training data—roughly 48 x 106 pixels—to get started.

This analysis has revealed that almost all of the CNN designs currently in use for the semantic segmentation of multispectral pictures are based on 1D or 2D architectures. The context throughout the height and width of the slice can be used by the 2D convolutional kernels to create predictions. However, as 2D CNNs only accept one slice as input, they are unable to use the context of subsequent slices by default. For the prediction of segmentation maps, voxel information from neighboring slices may be helpful. This problem is addressed by 3D CNNs, which use 3D convolutional kernels to predict segmentation for a volumetric area of an image. Performance can be enhanced by using interslice context effectively [14]. In this study work, semantic segmentation of a multispectral picture with seven channels and more extra feature classes was performed using modern end-to-end DCNN segmentation frameworks via a 3D U-Net convolutional neural network. It is believed that these methods will contribute to the creation of frameworks that are more discriminating and produce greater performance.

III. MATERIALS AND METHOD

In contrast to the methodologies previously discussed, the innovative 3D CNN architecture proposed in this study simultaneously processes the spatial and spectral components with genuine 3D convolutions, resulting in better use of the limited sample sets with fewer trainable parameters. According to this idea, the issue can be broken down into processing a number of volumetric representations of the image. As a result, each pixel is connected to n spatial neighborhoods and f spectral bands. Each pixel is therefore handled as an n×n×f volume. This architecture's key idea is to combine the conventional CNN network with a modification that applies 3D convolution operations rather than 1D convolution operators that only examine the spectral content of the data To achieve deep accurate representations of the image, various CNN layer blocks are stacked on top of one another. To start, a series of layers based on 3D convolution is introduced to deal with the three-dimensional input voxels. Each of these layers consists of a number of volumetric kernels that perform convolutions on the input's width, height, and depth axes all at once. A series of Fully Connected layers and a set of 1 1 convolution (1D) layers that ignore the spatial neighborhood follow such a stack of 3D convolutions. The suggested architecture essentially takes 3D voxels as input data and creates 3D feature maps first, which are then gradually reduced into 1D feature vectors across the layers. By selecting particular strides and padding configurations for the convolution filter, this process is ensured.

A. Proposed Framework

There are altogether 19069955 parameters in the architecture. By expanding the number of channels even before max pooling, we can prevent bottlenecks. This plan is used by us in the synthesis path as well. A N x N x F voxel tile of the image with 7 channels serves as the network's input. We have N x N x F voxels in the final layer, arranged in the x, y, and z directions, respectively. The approximate receptive field for each voxel in the expected segmentation is 155x155x180xm3 with a voxel size of 1:76x1:76x2:04xm3. As a result, each output voxel has access to enough information for effective learning.

ISSN No:-2456-2165

Prior to each ReLU, we also introduce batch normalization (BN). During training, each batch is normalized using its mean and standard deviation, and using these values, global statistics are updated. A layer that explicitly teaches scale and bias is added after that. These computed global statistics, along with the scale and bias that have been learned, are used to normalize data at test time. However, we only have one batch and a small number of samples. Utilizing the most recent statistics at test time is ideal in these applications. The weighted SoftMax loss function is a crucial component of the architecture that enables us to train on sparse annotations. It is possible to learn from only the named pixels and, therefore, generalize by setting the weights of the unlabeled pixels to zero.

B. Model Training

In this study, a different 3D U-Net network will be used to produce the 3D U-Net Layer. The initial set of convolutional layers in U-Net is broken up by max-pooling layers, gradually lowering the input image's resolution. These layers are preceded by a string of convolutional layers with up-sampling operators interleaved, gradually raising the input image's resolution. The network can be drawn with a symmetric shape similar to the letter U, hence the name U-Net. In order to make the input and output of the convolutions equal in size, this research alters the U-Net to use zero-padding in the convolutions. So, using a few pre-selected hyper-parameters, we build a U-Net.

C. Data Source and Choice of Metrics

In this study, the network is trained using a high-resolution multispectral data set. A drone was used to take the image collection over the research area in Gombe Metropolis. There are nine object class labels on the data's training, validation, and test sets. The data file is 3.0 GB in size. In this study, accuracy is used to assess how well the proposed model performs. This performance metric, which deals with the model's correct predictions, is expressed as:

$$\text{Accuracy} = \frac{TP+TN}{TP+FP+FN+TN} \quad (1)$$

$$\text{Precision} = \frac{TP}{TP+FP} \quad (2)$$

$$\text{Recall} = \frac{TP}{TP+FN} \quad (3)$$

IV. IMPELEMTATION OF THE SYSTEM

MATLAB was used for our experiment. This study trains the network using a high-resolution multispectral data collection and then implements the suggested model. Over the study area in the Gombe metropolis, a drone was used to take the image collection. As depicted in Fig. 1, the data includes labeled training, validation, and test sets with 9-item class labels. The data file is 3.0 GB in size.

The experiment was carried out using MATLAB 2021a, a 64-bit version of Microsoft Windows 10, 8GB of RAM, and an Intel(R) Core(TM) i7-4000M @ 2.4 GHz processor. This study trains the network using high-resolution multispectral data in order to execute the suggested model.

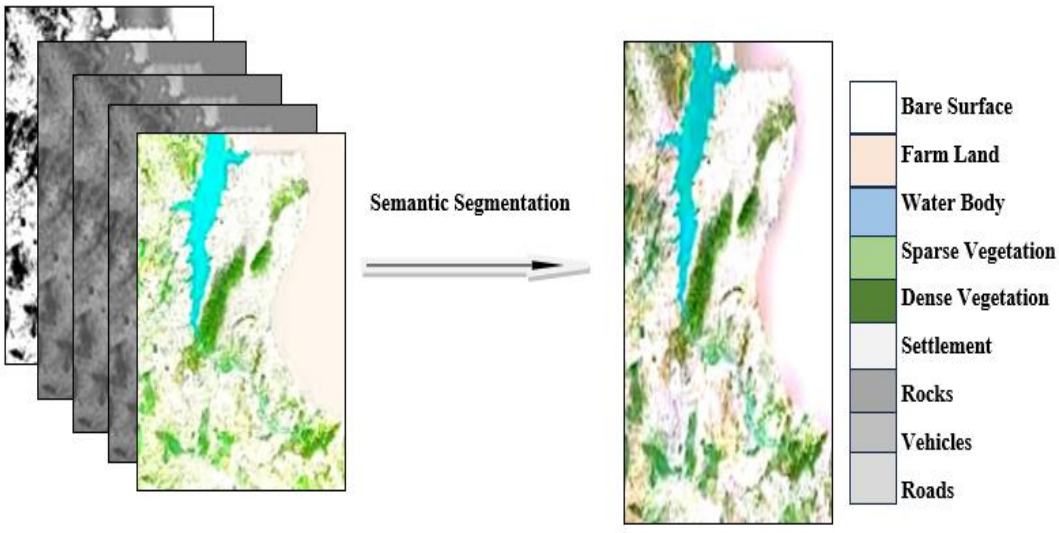

Fig. 1: Multispectral image with 9 object class labels

The multispectral picture data is organized as arrays with dimensions of numChannels-by-width-by-height. to modify the data so that the third-dimensional channels are used. The third, second, and first picture channels in Figure 2 are the RGB color channels. given the color element of the training, validation, and test images as a montage.

Fig. 2: 3 RGB component of training image (left), validation image (center) and test image (right)

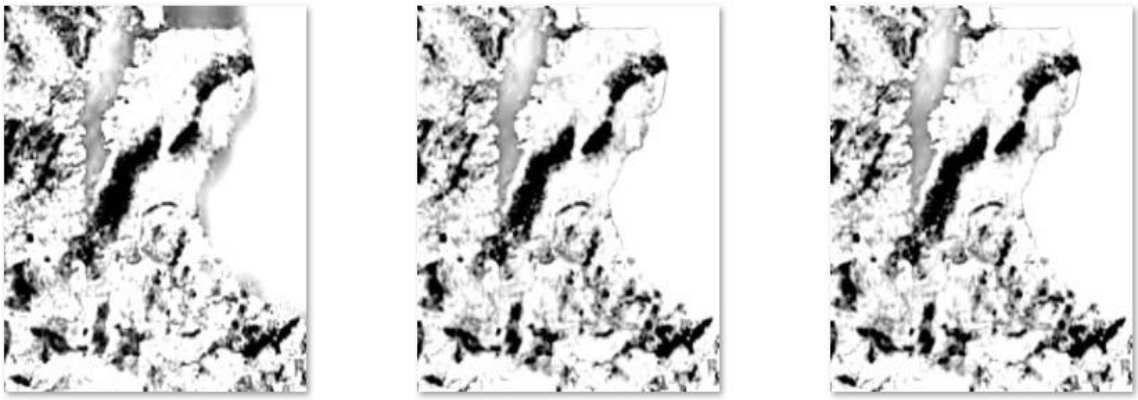

Fig. 3: IR channel 1 (left), 2 (center) and test image (right)

Fig. 4: Mask of training image (left), validation image (center) and test image (right)

The segmentation's ground truth data is contained in the labeled images, where each pixel has been allocated to one of the classes. The 9 classes and their IDs are listed in Table 1.

Table 1: Image Classes and IDs for THE DATASETS

IDs	Class Name
0.	Other Class/Image Border
1.	Road
2.	Vehicle (Car, Truck, or Bus)
3.	Rocks
4.	Settlement
5.	Dense Vegetation
6.	Sparse vegetation
7.	Water Body
8.	Farm land
9.	Bare surface

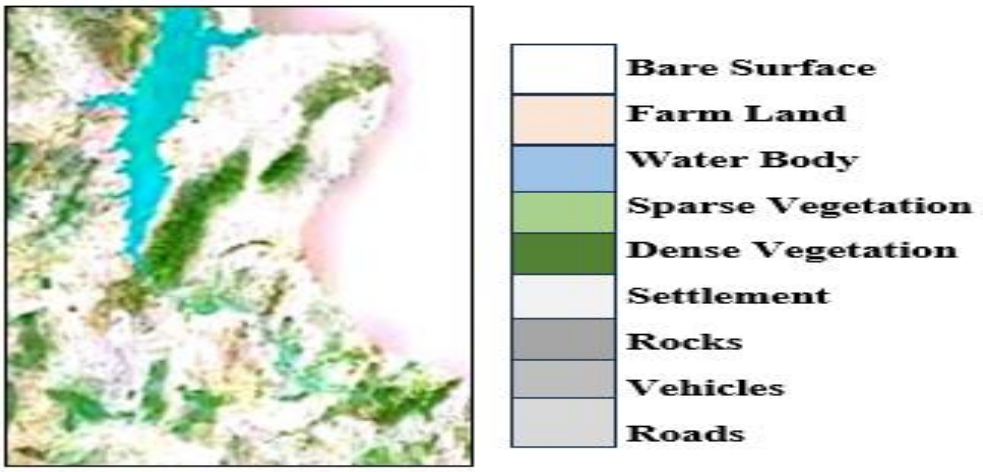

Fig. 5: Training labels from the datasets

By dividing the total number of vegetation pixels by the total number of valid pixels, this work's final objective is to determine the extent of plant cover in the multispectral image. Table 2 shows the settings for the hyper parameters and training options.

Table 2: Image Classes and IDs for THE DATASETS

Parameters	**Settings**
Initial Learning Rate	0.05
Max Epochs	150
Mini batch Size	16
l2reg	0.0001
Momentum	0.9
Learn Rate Schedule	Piecewise
Shuffle	every-epoch
Gradient Threshold Method	l2norm
Gradient Threshold	0.05
Verbose Frequency	20

As a result, we can now semantically segment the multispectral image using the constructed 3D-U-Net. The section above presents the prediction outcomes in terms of accuracy, precision, and recall.

A. Results

Segmentation of image patches is carried out utilizing the semantics function during the forward pass on the trained network. Figure 6 displays a representative sample of the segmented image.

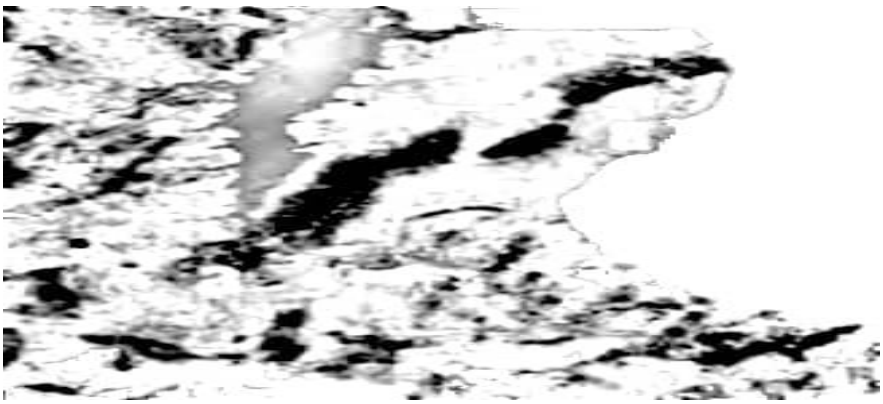

Fig. 6: Segmented Image.

The baseline reality labels and the segmented image are both stored as PNG files. As illustrated in Fig. 7, they were utilized to calculate accuracy measures by superimposing the segmented image over the RGB validation image with an equalized histogram.

ISSN No:-2456-2165

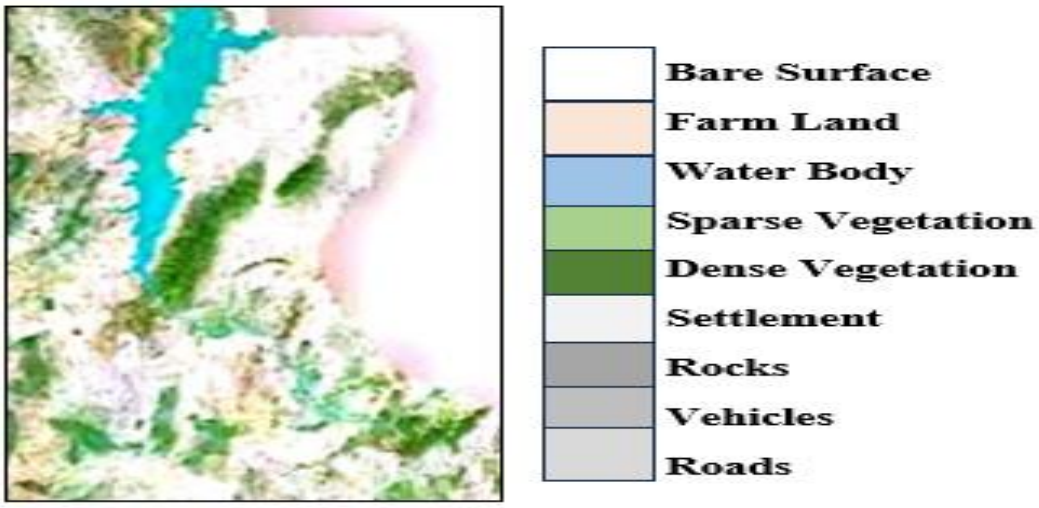

Fig. 7: label Validation Image

A training accuracy assessment was conducted after 100 iterations of the model, as shown in Table 3. A 90.698% accuracy was achieved after training. Generally, the better the model, the higher the accuracy of categorization.

By dividing the total number of vegetation pixels by the total number of valid pixels, this study aims to determine how much vegetation is present in the multispectral image. As shown in the segmented image, the overall plant coverage for the present study is 24.42%.

B. Validation

In this section, we evaluate the performance of the proposed model against other classification framework using the collected datasets. Table 3 depicts the mean-class accuracy (AA) on the test set compare with other existing approaches.

Table 3: Performance comparison against classical approaches

Model	Mean Accuracy (%)	Precision	Recall
Proposed	91	95	97
MLP	31	41	50
KNN	28	32	29
SVM	30	39	31
Sharp Mask	58	68	61
Refine-Net	60	59	66

Table 3 makes it very clear that in terms of accuracy, precision, and recall, the proposed deep learning algorithms surpass the traditional technique. For a clear understanding of how the difference in performance trend continues, this conclusion is further illustrated in Fig. 8 in a graphical depiction.

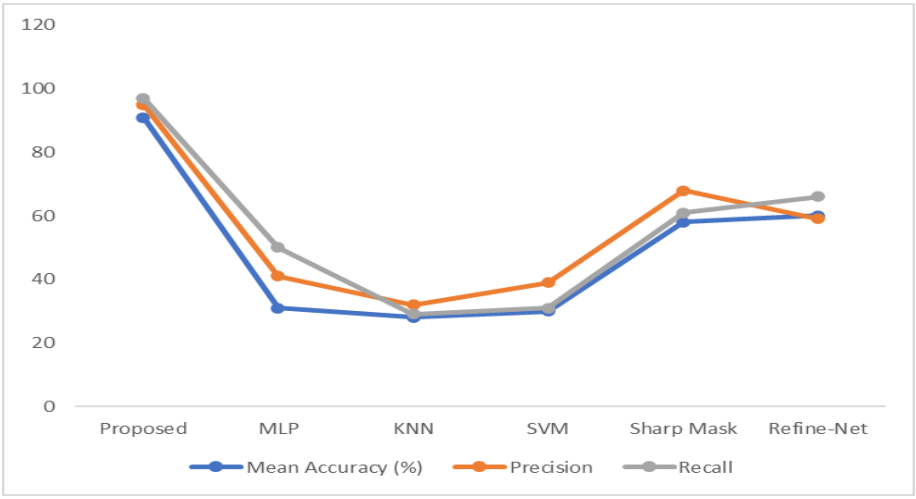

Fig. 8: Performance Comparison with classical approaches

ISSN No:-2456-2165

V. CONCLUSION

As seen in the in-classification task, a precision and recall value within the range of 91% – 100% can be concluded as an excellent performance that can support the accuracy score by a given classification model. Thus, From table 5. It can be notices that the proposed model achieved precision value of 95% and recall value of 97%. generally, we can conclude that the proposed model has significantly improve the classification performance other existing approaches used in this context. This study has demonstrated that semantic segmentation of a multispectral image with seven channels can be performed using newer end-to-end DCNN segmentation frameworks via a 3D U-Net convolutional neural network, which can address the majority of the difficulty in DL for RS aspects. It is thought that these methods can assist in the creation of frameworks with greater discrimination and better performance. In contrast to other research, this study also determines the segmented image's vegetation cover. The findings of this study demonstrate that this upgraded model has greater potential for image processing in the context of remote sensing and the enhancement of satellite images.

ACKNOWLEDGMENT

This study was supported by the Tertiary Education Trust Fund (TET Fund) Institutional Based Research (IBR) Fund for Federal Polytechnic Kaltungo, Gombe state. Nigeria.

REFERENCES

[1.] Yuan, X., J. Shi, and L. Gu, A Review of Deep Learning Methods for Semantic Segmentation of Remote Sensing Imagery. Expert Systems with Applications, 2020: p. 114417.

[2.] Girshick, R., et al. Rich feature hierarchies for accurate object detection and semantic segmentation. in Proceedings of the IEEE conference on computer vision and pattern recognition. 2014.

[3.] Wu, Z., et al., Semantic segmentation of high-resolution remote sensing images using fully convolutional network with adaptive threshold. Connection Science, 2019. **31**(2): p. 169-184.

[4.] Abdel-Hamid, O., et al. Applying convolutional neural networks concepts to hybrid NN-HMM model for speech recognition. in 2012 IEEE international conference on Acoustics, speech and signal processing (ICASSP). 2012. IEEE.

[5.] Praveena, S. and S. Singh. Hybrid clusteing algorithm and Neural Network classifier for satellite image classification. in 2015 International Conference on Industrial Instrumentation and Control (ICIC). 2015. IEEE.

[6.] Shuang, W., et al., *A deep learning framework for remote sensing image registration.* ISPRS Journal of Photogrammetry and Remote Sensing, 2018. **145**(1): p. 148-164.

[7.] Jony, R.I., et al. Ensemble Classification Technique for Water Detection in Satellite Images. in 2018 Digital Image Computing: Techniques and Applications (DICTA). 2018. IEEE.

[8.] Wang, L., et al., Object-based convolutional neural networks for cloud and snow detection in high-resolution multispectral imagers. Water, 2018. **10**(11): p. 1666.

[9.] Kemker, R., C. Salvaggio, and C. Kanan, *Algorithms for semantic segmentation of multispectral remote sensing imagery using deep learning.* ISPRS journal of photogrammetry and remote sensing, 2018. **145**: p. 60-77.

[10.] Wurm, M., et al., Semantic segmentation of slums in satellite images using transfer learning on fully convolutional neural networks. ISPRS journal of photogrammetry and remote sensing, 2019. **150**: p. 59-69.

[11.] Mi, L. and Z. Chen, *Superpixel-enhanced deep neural forest for remote sensing image semantic segmentation.* ISPRS Journal of Photogrammetry and Remote Sensing, 2020. **159**: p. 140-152.

[12.] You, J., W. Liu, and J. Lee, *A DNN-based semantic segmentation for detecting weed and crop.* Computers and Electronics in Agriculture, 2020. **178**: p. 105750.

[13.] Bhatnagar, S., L. Gill, and B. Ghosh, Drone image segmentation using machine and deep learning for mapping raised bog vegetation communities. Remote Sensing, 2020. **12**(16): p. 2602.

[14.] Hamida, A.B., et al. Deep learning approach for remote sensing image analysis. in Big Data from Space (BiDS'16). 2016. Publications Office of the European Union.