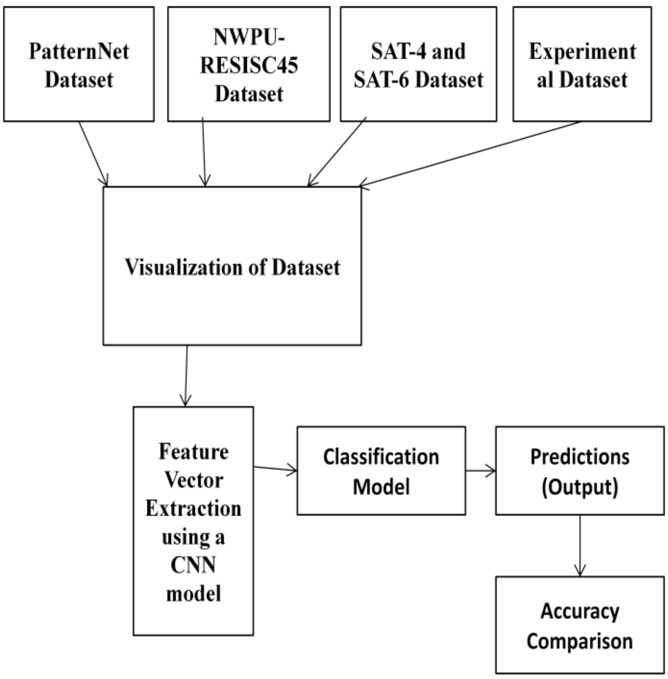
MOL2NET, International Conference Series on Multidisciplinary Sciences
USEDAT-08: USA-Europe Data Analysis Training Program Workshop,
UPV/EHU, Bilbao-MDC, Miami, USA, 2020

Evaluation of Datasets for CNN based Image Classification

Shraddha Marbhal [a], *Minakshi Kumar* [b]

[a] *Department of Technology SPPU Pune Maharashtra India*
[b] *Indian Institute of Remote Sensing ISRO Dehradun Uttarakhand India*

Graphical Abstract



Abstract

In this paper, we address the challenge of land use and land cover classification (LULC) using Convolution Neural Networks (CNN) on existing remote sensing datasets and compare the obtained results in an Indian Urban area context. This paper showcases the theoretical and experimental study of various large-scale, high-resolution remote sensing datasets. An image retrieval dataset is used to perform image classification and promising results are found. Also, a small-scale dataset is used and modified from a high-resolution large-scale data as per the requirement. Different datasets with different dimensions and spectral bands are used for the study. The results and comparisons of various datasets are tabulated. From a dataset point of view, classification or categorization techniques can be developed and assessed by making use of image retrieval datasets but this doesn't work the other way around.

This paper also provides the literature with standard outcomes for future research on datasets for Machine Learning based image classification especially in terms of reducing memory consumption of computers and fastening the process of execution. The resulting

	classification system finds its use in a large number of Earth observation applications.

Introduction

Advancements in Artificial Intelligence have made the autonomous large-scale image analysis of imagery possible. Also advancements in remote sensing have encountered drastic changes in the expanded spatial resolution of the imagery just as the expanded rate of acquisition. These alterations surely affect the way remote sensing images have had been used and handled. The increasing spatial resolution gives new chances to propelling remote sensing image investigation and comprehension, making it more feasible to create novel methodologies that seem impractical previously. The expanded rate of acquisition empowers to get an impressive volume of remote sensing information regularly. Yet, it triggers a question on efficient management of the large data collections, so that the data of interest can be accessed quickly.

In image classification, one or more semantic labels are assigned to an image whereas in image retrieval a target set that is similar to a query image is used to recognize images. Classification is generally performed using a classifier that is trained using a set of labeled images. On the other hand, image retrieval from any dataset is achieved by comparing features drawn-out from the query image to features drawn-out from the target images. These correlations are used to rank the target images in order of declining similarity. From a Dataset point of view, Classification techniques can be developed and assessed by making use of image retrieval Datasets but this doesn't work the other way around. The images in image retrieval should not divert scenes or backgrounds.

This paper showcases theoretical and experimental study of various large-scale, high resolution remote sensing datasets. An image retrieval dataset is used to perform image classification and promising results are found. Also a small-scale dataset is used and modified from a high resolution large-scale data as per the requirement. This provides the literature with standard outcomes for future research on datasets for Machine Learning based image classification. The results and comparisons of various datasets are shown in Table 1.

Although there are many datasets available for classification of satellite images but as the study is related to Machine Learning based classification thus only scene-level classification and its related datasets are reviewed and studied in detail. It was decided to evaluate the performance of these datasets for classifying an independent study area in a part of city in Northern India using CNN.

Materials and Methods

Two existing three-band datasets and two four-band datasets are reviewed in this paper. Also, from large satellite imagery one high-resolution dataset was extracted and modified as per the requirement. A detailed review of all the datasets used is mentioned below.

PatternNet Dataset: It's one of the largest freely accessible high-resolution benchmark dataset created by researchers specifically, for remote sensing image retrieval (RSIR) application. It is consisting of 38 scene classes which have covered a large number of land-forms. Every individual class holds 800 images evaluating 256×256 pixels. The images in PatternNet are of US urban areas, gathered from Google Earth imagery by means of the Google Map API by its creators. A large number of images make it more reliable and acceptable for deep learning-based RSIR approaches. The sample images visualized from PatternNet dataset are shown in Figure 1.

NWPU-RESISC45 Dataset: It's one of the uninhibitedly accessible and open datasets which comprises of 31,500 remotely sensed images which are partitioned into a total of 45 scene groups. Every individual scene group has 700 images with a fixed dimension of 256×256 pixels in the form of Red, Green, and Blue (RGB) concealing groups. The spatial resolutions fluctuate in the range of 30m to 0.2m per pixel for a large portion of the scene classes with the exception of the classes of the mountain, island, lake, and snow-berg that have even lower spatial resolutions. The dataset was extricated in the field of remote sensing image translation, from Google Earth (Google Inc.) that maps the Earth by the superimposition of images procured from satellite imageries, airborne photography, and the Geographic Information System (GIS) onto a 3D globe. A total of 31,500 remote detecting images have covered several nations and regions throughout the world. Figure 2 represents sample images visualized from NWPU-RESISC45 dataset.

Four bands (RGB NIR): SAT-4 Dataset, SAT-6 Dataset

SAT-4 is a direct open dataset including 500,000 image patches where each image fix is size normalized to 28×28 pixels covering four far-reaching land cover classes- barren land, trees, prairie, and a class that involves all land-cover classes other than the three referenced. Around 400,000 patches (containing four-fifths of the total dataset were picked for the training and the rest 100,000 (one-fifths) were picked as the testing dataset. It is ensured that the training and testing datasets have a spot with a disjoint plan of satellite image tiles. Figure 3 shows contents of SAT-4 dataset.

SAT-6 is genuinely similar to the SAT-4 dataset, a clearly open dataset involving a total of 405,000 image tiles every last one of dimension 28×28 and covering six land-cover classes-barren lands, streets, structures, trees, field, and water-bodies. A total of 324,000 images (including four-fifths of the whole dataset) was picked as the training dataset and a sum of 81,000 (one-fifth) was picked as the testing dataset. Like SAT-4, from disjoint NAIP tiles, the train and test sets were chosen by its researchers. Figure 4 shows contents of SAT-6 dataset.

For both SAT-4 and SAT-6 datasets a dimension of 28×28 is chosen as the window size to keep up a significantly bigger context, and simultaneously not to make it as large as to drop the relative statistical properties of the target class conditional distributions within the relevant window. It fairly avoids interclass overlaps within a selected and labeled image patch.

Modified Experimental Dataset: A part of the Chandigarh area situated in northern part of India was selected from a high-resolution satellite imagery having 0.5m resolution. This large patch of imagery was further cut into equal size of 256×256 which gave 90 images out of which only 20 images were used for experimental analysis and classification. All the images were further converted into JPEG format having three-bands (RGB) only.

A CNN model was developed using MobileNetV2 architecture, and model scene-level classification was carried out and outputs were obtained into semantic categories with their respective probabilities (ranging between 0 and 1).

Results and Discussion

Dataset	Resolution (m)	Bands	Size	Classes	Images
PattenNet	4.693- 0.062	RGB	256×256	38	30400
NWPU-RESISC45	30- 0.2	RGB	256×256	45	31500
SAT 4	1	RGB-NIR	28×28	4	500000
SAT 6	1	RGB-NIR	28×28	6	405000
Modified Experimental Dataset	0.5	RGB	256×256	5	90

Table 1: Comparison of Datasets

It is observed that one out of four classes in SAT-4 dataset is labeled as ‘none’, which signifies that it contains all classes other then the three existing classes (barren land, grass land, trees) in the dataset.

But to verify classes in SAT-4 and SAT-6 dataset visually, pixel level visualization should be used because both the datasets has images of size 28 × 28 which makes it difficult to recognize visually. On other hand, the rest of the datasets used are of size 256 × 256 in RGB formats thus are easy to recognize.

With machine learning based image classification, scene level classification is better as it gives numerous points of interest when contrasted with pixel level or object level classification. Firstly, the successful accomplishment of any Machine Learning-based image classification problem relies on the size and quality of the training dataset. In this regard, it is generally quicker and less complex to assemble a training dataset for a scene-based classification of images when contrasted with the other types of image classification.

Besides, classifying obscure classes (such as cloud shades, haze, and so forth) in satellite or aerial image data, both the object-level and the pixel-level convey minimal semantic meanings, as their properties could be confused with other land cover types. For example, significantly influenced by background features, cloud shadows signal are sometimes misunderstood as water and or dark vegetation signal. At last, a scene-level characterization will in general have higher speculation capacity, as the logical data of the entire scene is considered in the learning procedure.

All the datasets mentioned were analyzed and visualized. Refer to Figure below for the sample visuals of PatternNet dataset.

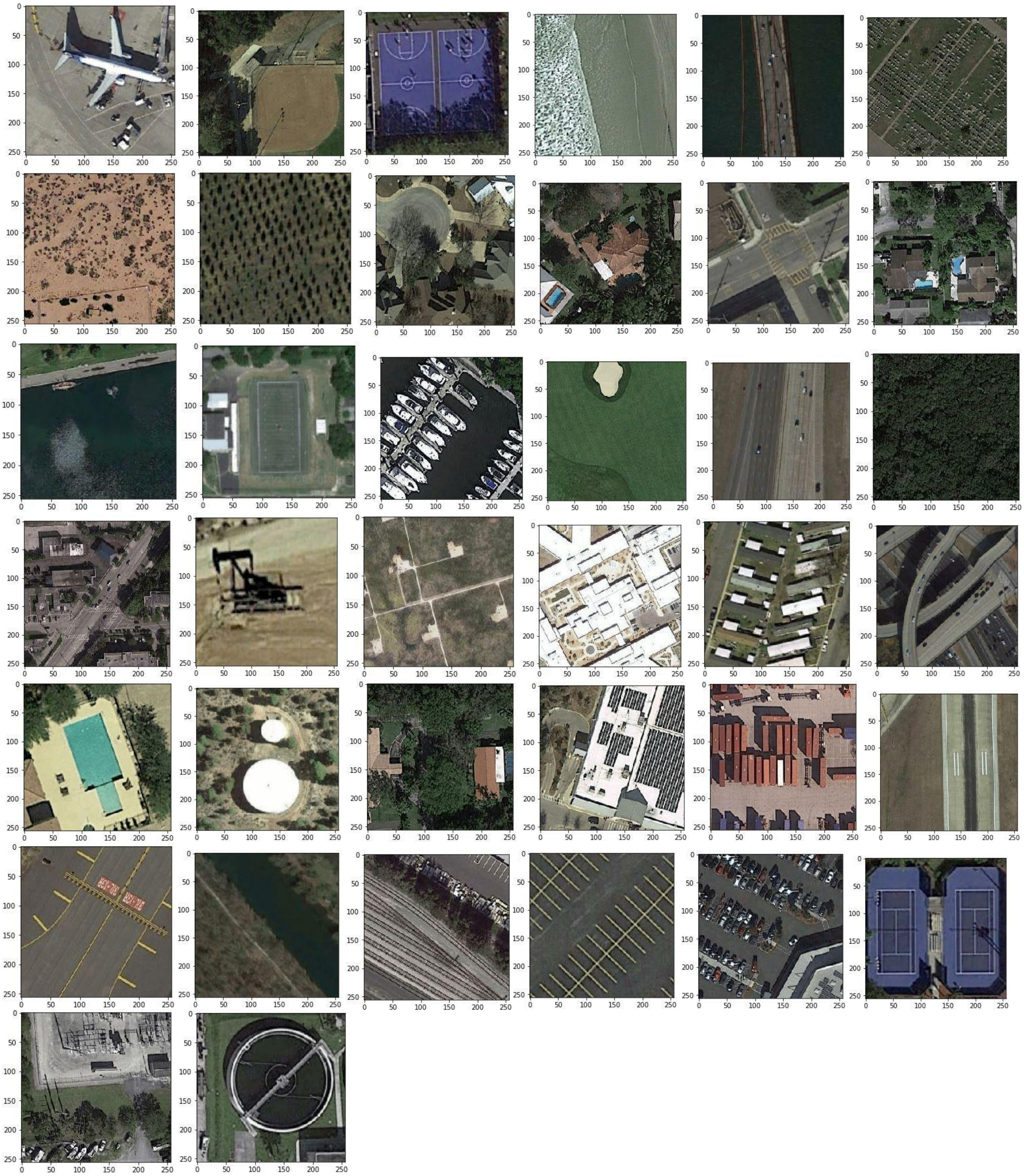


Figure1: Visualized sample images from PatternNet dataset

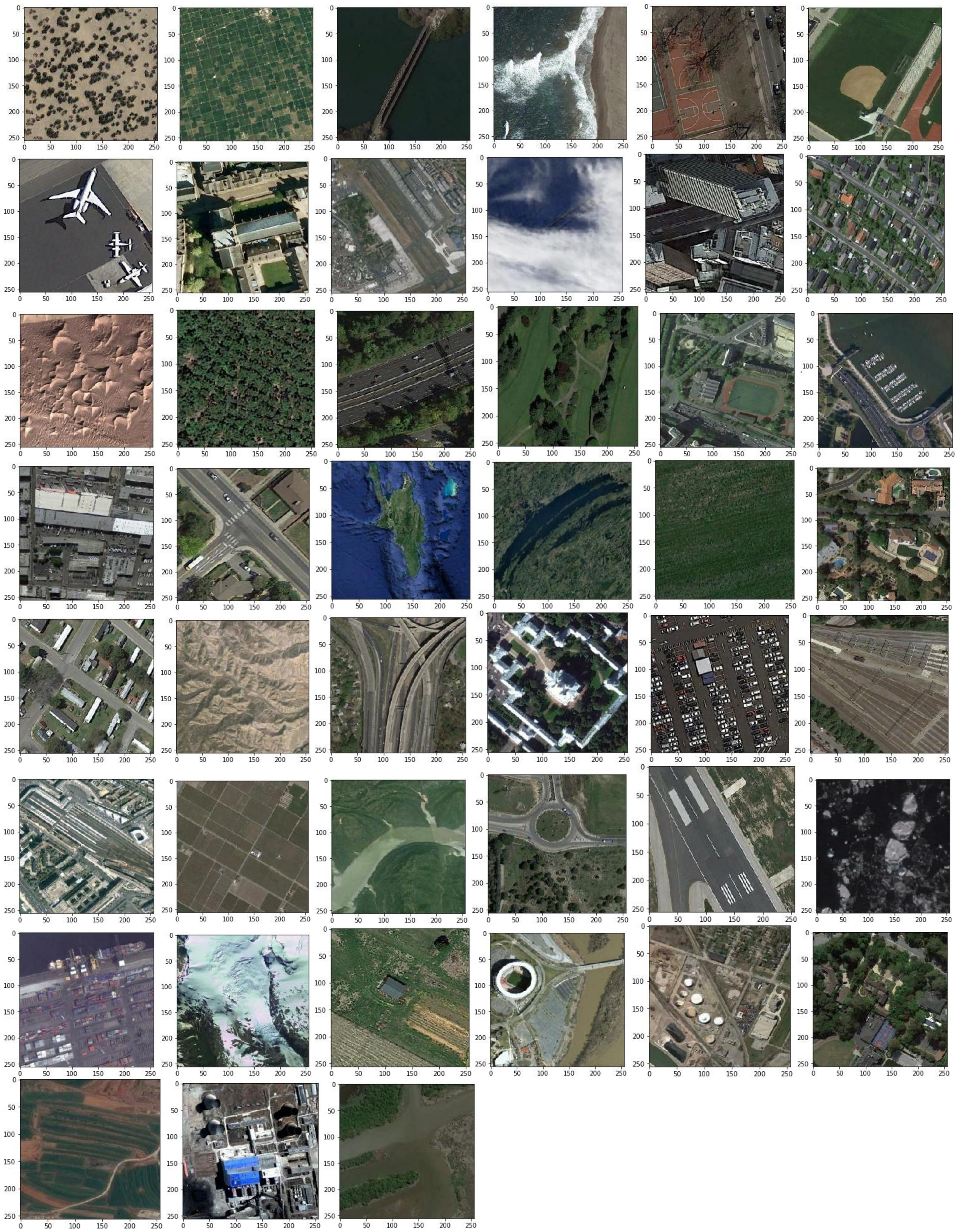


Figure 2: NWPU-RESISC45 dataset visualization

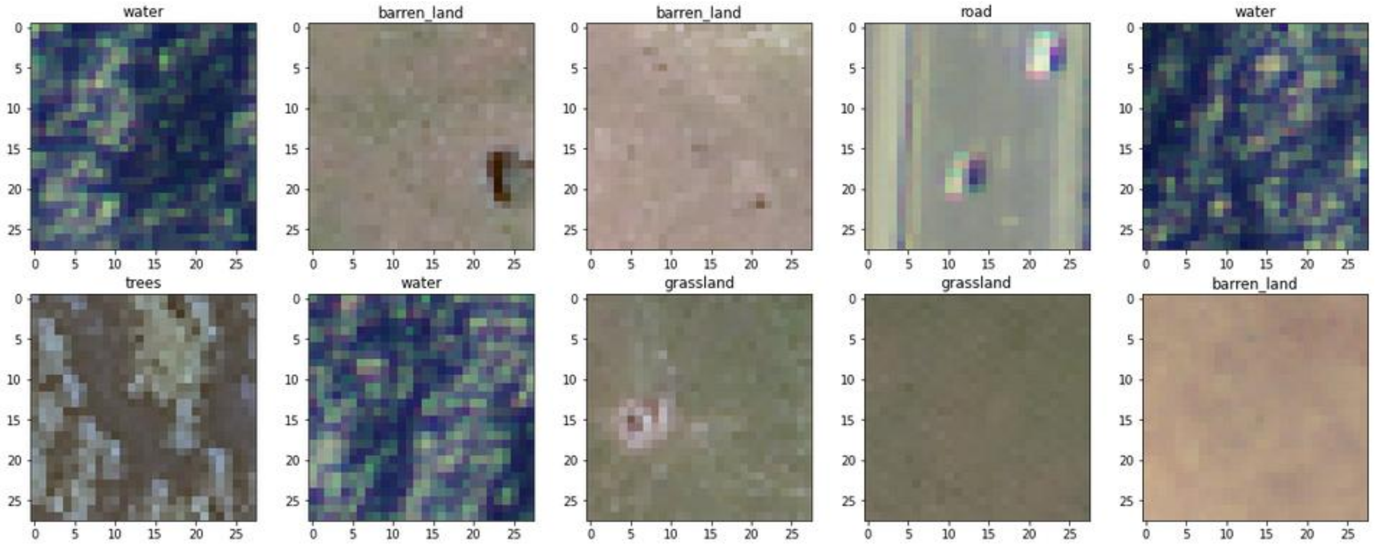


Figure 3: Sample visualized data from SAT-6 datasets.

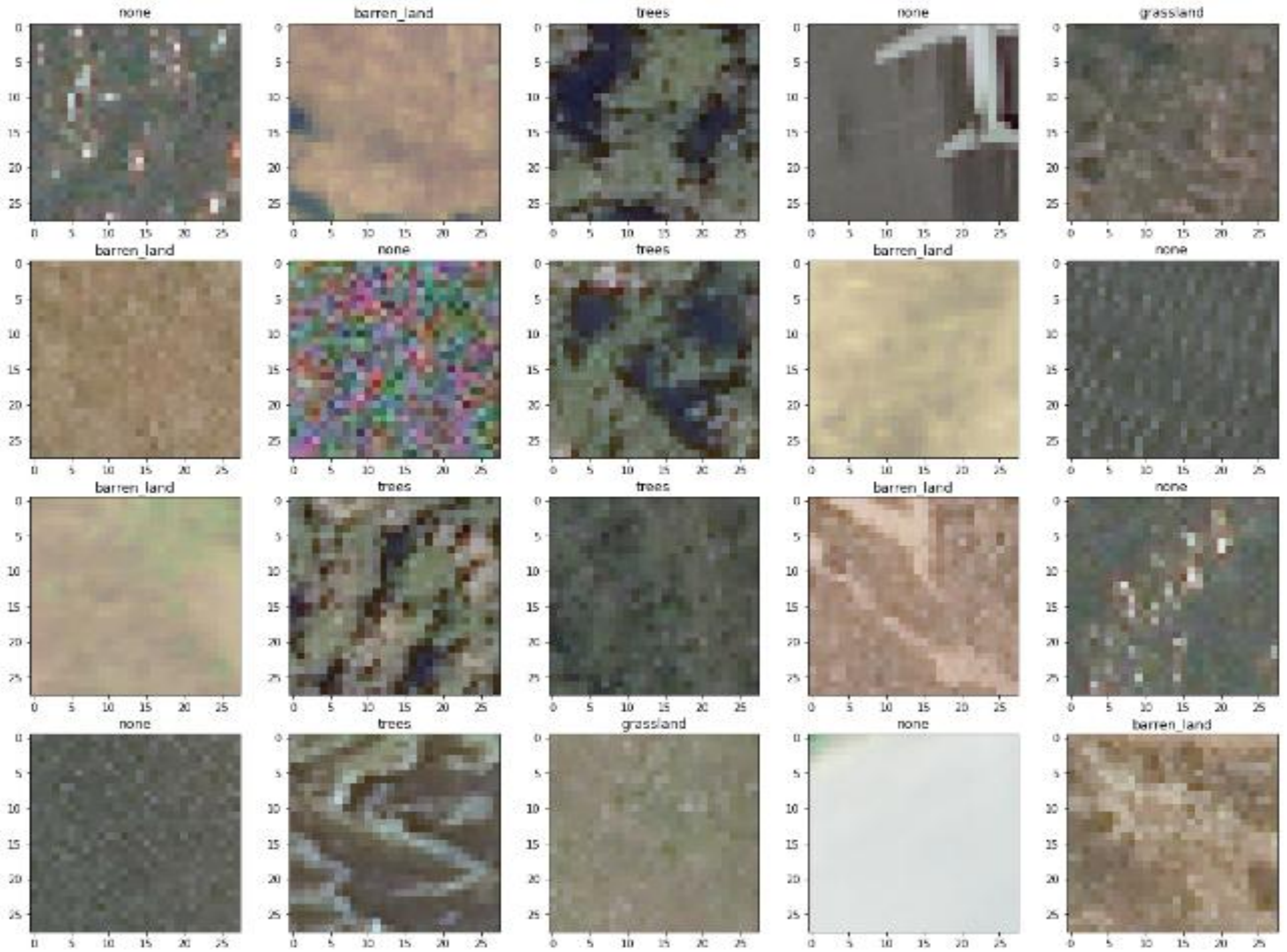


Figure 4: Sample data visualized from SAT-4 dataset

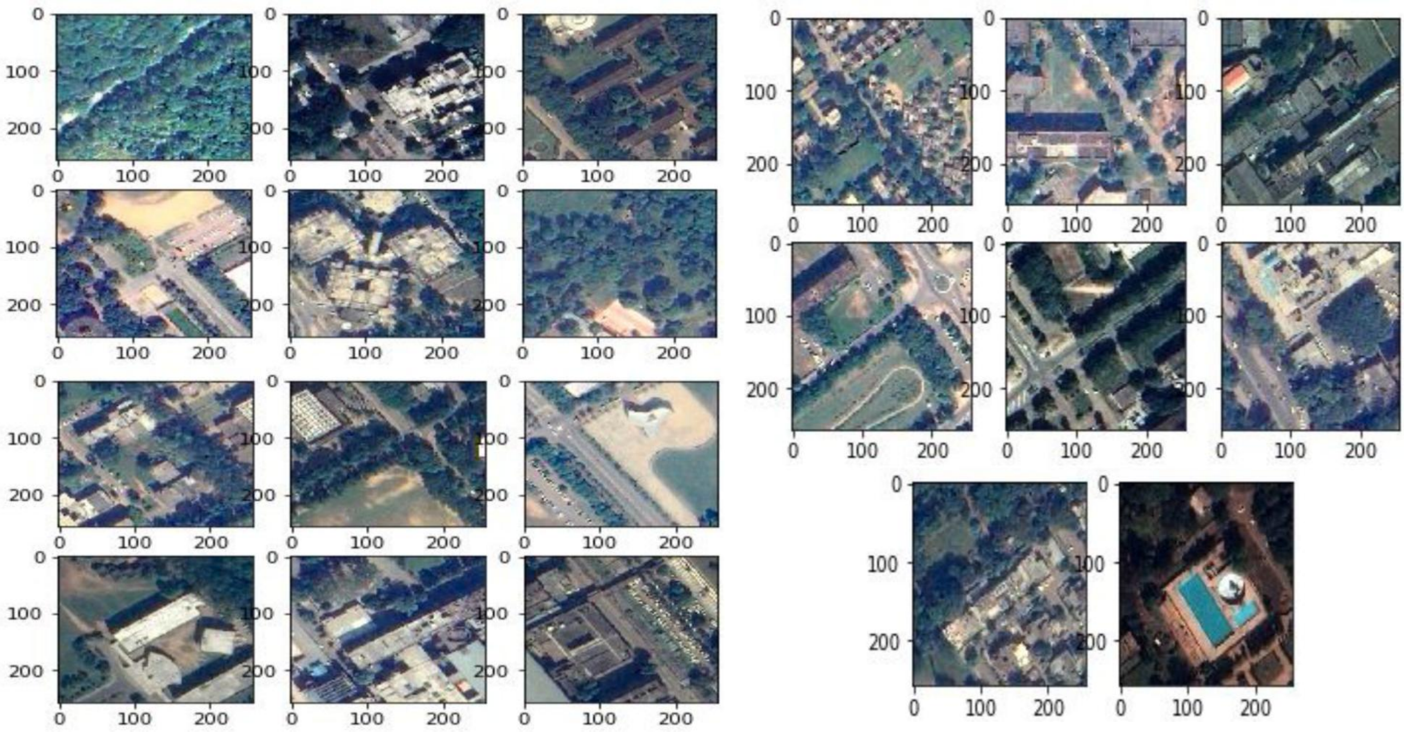
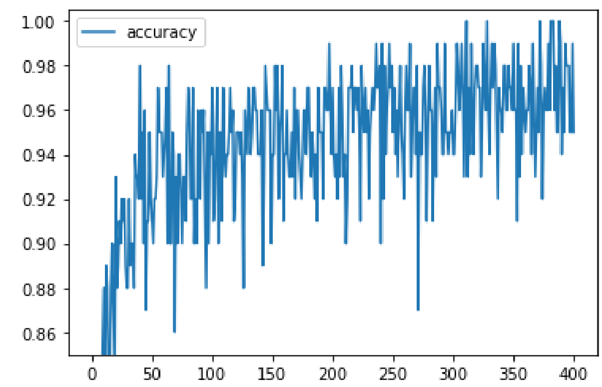


Figure 4: Sample data visualized from Modified Experimental Dataset

```
#plt.plot(train_accuracy['Value'], label='train')
plt.plot(validation_accuracy['Value'], label='accuracy')
plt.ylim((0.85, 1.005))
plt.legend();
```



Validation accuracy is about 96%

Figure 5: Validation Accuracy Plot

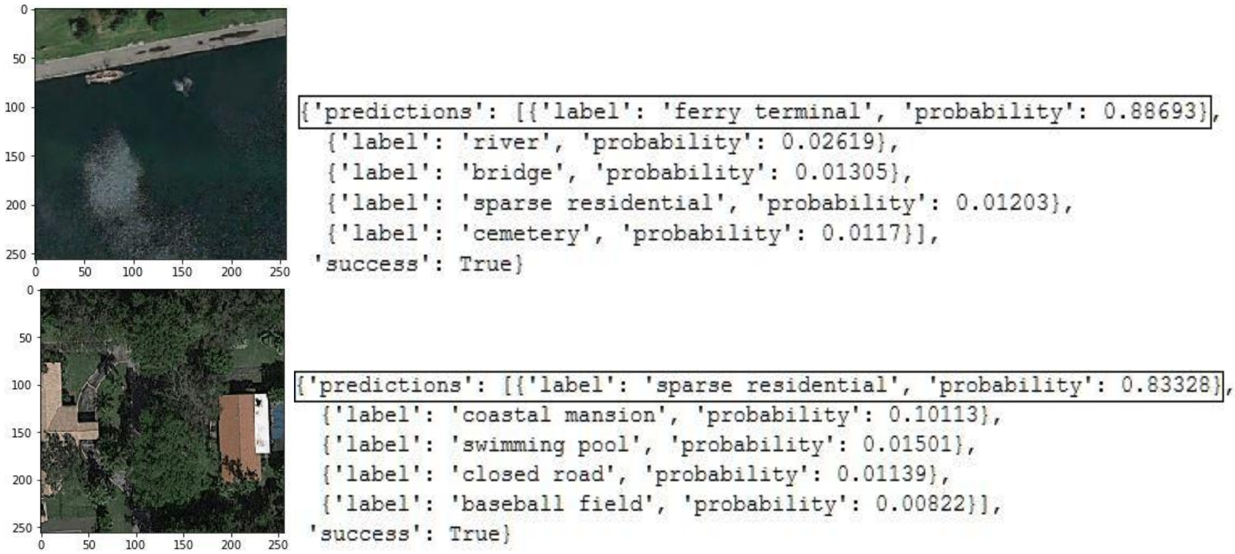


Figure 6: Sample predictions from PatternNet Dataset

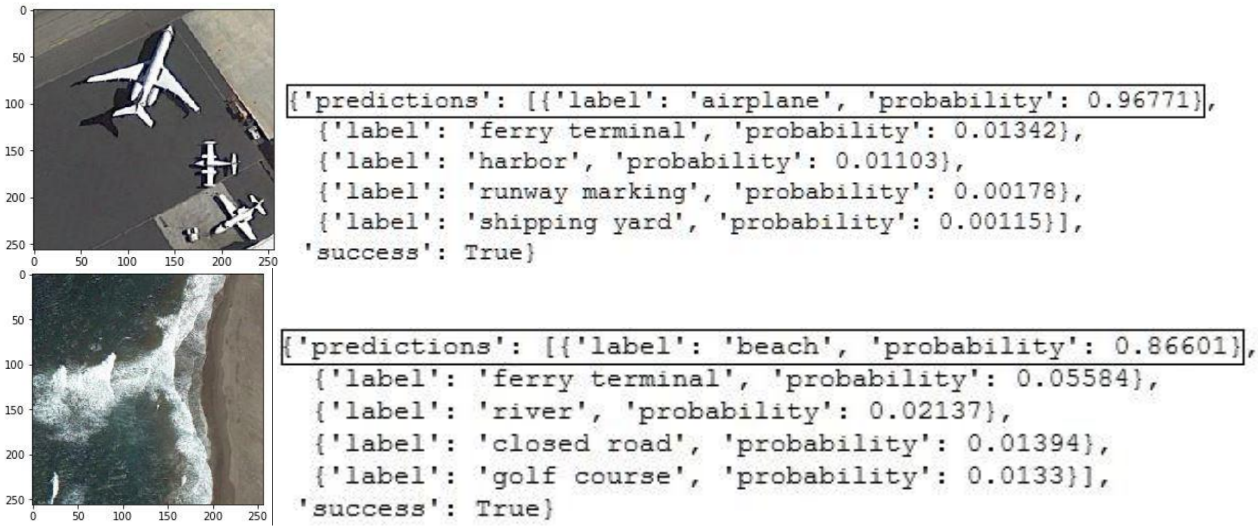


Figure 7: Sample predictions from NWPU-RESISC45 dataset

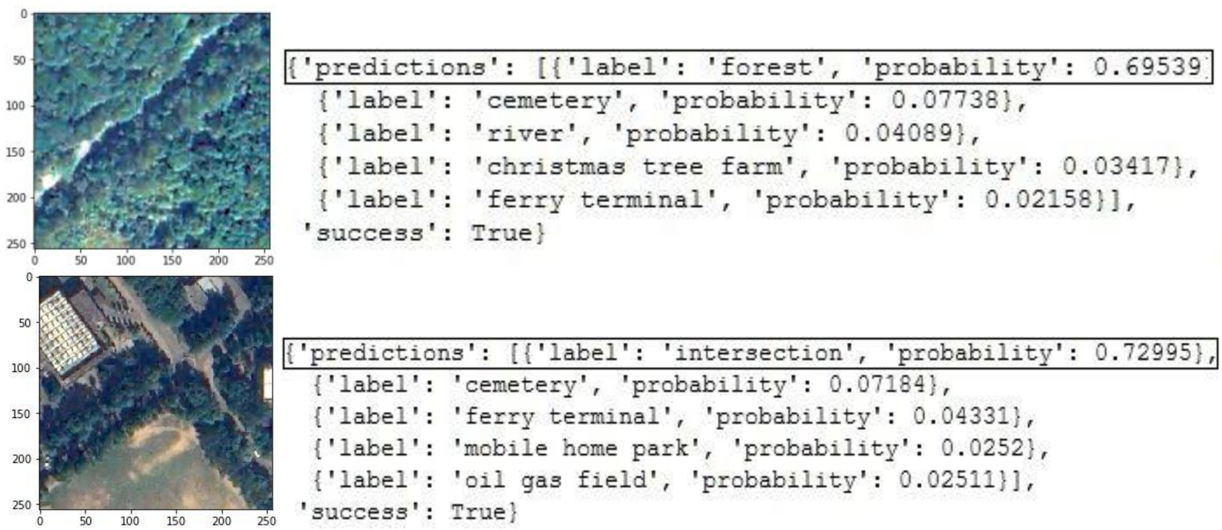


Figure 8: Sample predictions from Modified Experimental Dataset

The modified experimental dataset was trimmed to contain only RGB band and image scenes of 256 × 256 pixels (i.e. 800 × 800 m) in size. It is also observed that each of the scene images of all the datasets used has multiple labels with corresponding probability ranging between 0 and 1, such that the total sum of all the probabilities is 1.

A notable limitation of using scene level approach for image classification is that there is loss of information because of the existence of class within a scene without identifying its inch-perfect location which can be overcome by translating it to pixel level one by convincing CNN during training to add further weights on the pixels.

Conclusions

It is observed that four bands (R, G, B, and NIR) imagery depicts a different type of detail in the imagery than natural color, and the vegetative and cropping boundaries are more distinct in any image. Yield health and soil moisture variations are more apparent with RGB-NIR bands.

Four band imagery is more reliable than three bands (RGB) in agriculture because a bright red color is an indication of healthy vegetation. Variations in the red color indicate stressed vegetation. These stresses can incorporate an absence of fruitfulness, bug pervasion, soil insufficiencies, and over or under-watering. The unique spectral reflectance of vegetation additionally gives significant data that can be utilized to develop land-classifications.

As we know that artificial satellite gives data in various numbers of bands, usually 4 to 12 bands. For satellite image classification using machine learning techniques it is recommended to use necessary bands only and drop-out other bands in order to minimize consumption of overall memory utilized by computer. Thus three bands or four bands (for better classification accuracy) datasets should be used and developed from high resolution large satellite imagery for the application of image classification using machine learning.

References

1. Shraddha Marbhal, Minakshi Kumar, Land Spread Characterization of Satellite Pictures Utilizing Deep Learning Techniques, https://sciforum.net/paper/view/conference/6871, (doi: 10.3390/mol2net-06-06871), July 2020.
2. Zhou, W., Newsam, S., Li, C., & Shao, Z. (2018). PatternNet: A benchmark dataset for performance evaluation of remote sensing image retrieval. ISPRS Journal of Photogrammetry and Remote Sensing, 145, 197-209.
3. G. Cheng, J. Han and X. Lu, "Remote Sensing Image Scene Classification: Benchmark and State of the Art," in Proceedings of the IEEE, vol. 105, no. 10, pp. 1865-1883, Oct. 2017, doi: 10.1109/JPROC.2017.2675998.
4. Saikat Basu, Sangram Ganguly, Supratik Mukhopadhyay, Robert Dibiano, Manohar Karki and Ramakrishna Nemani, DeepSat - A Learning framework for Satellite Imagery, ACM SIGSPATIAL 2015.
5. United States Department of Agriculture, Four Band Digital Imagery, INFORMATION SHEET, http://www.apfo.usda.gov, June2013

6. Tatem AJ, Goetz SJ, Hay SI. Fifty Years of Earth Observation Satellites: Views from above have lead to countless advances on the ground in both scientific knowledge and daily life. Am Sci. 2008; 96(5):390-398. doi:10.1511/2008.74.390
7. Tamouk, Jamshid & Lotfi, Nasser & Farmanbar, Mina. (2013). Satellite image classification methods and Landsat 5TM bands.
8. L. Gómez-Chova, D. Tuia, G. Moser and G. Camps-Valls, "Multimodal Classification of Remote Sensing Images: A Review and Future Directions," in Proceedings of the IEEE, vol. 103, no. 9, pp. 1560-1584, Sept. 2015, doi: 10.1109/JPROC.2015.2449668.
9. E. Piazza, "Minimum band set to classify AVHRR images," IGARSS 2001. Scanning the Present and Resolving the Future. Proceedings. IEEE 2001 International Geoscience and Remote Sensing Symposium (Cat. No.01CH37217), Sydney, NSW, Australia, 2001, pp. 1901-1903 vol.4, doi: 10.1109/IGARSS.2001.977110.
10. Gu, Y.; Wang, Y.; Li, Y. A Survey on Deep Learning-Driven Remote Sensing Image Scene Understanding: Scene Classification, Scene Retrieval and Scene-Guided Object Detection. Appl. Sci. 2019, 9, 2110
11. Fan, L.; Zhao, H.; Zhao, H. Distribution Consistency Loss for Large-Scale Remote Sensing Image Retrieval. Remote Sens. 2020, 12, 175.