IGARSS 2021 - 2021 IEEE International Geoscience and Remote Sensing Symposium | 978-1-6654-0369-6/21/$31.00 ©2021 IEEE | DOI: 10.1109/IGARSS47720.2021.9554290

SOIL TYPE CLASSIFICATION FROM HIGH RESOLUTION SATELLITE IMAGES WITH DEEP CNN

Abhinav Pandey, Devesh Kumar and Debarati B. Chakraborty
Dept. of Computer Science and Engineering,
Indian Institute of Technology Jodhpur, Jodhpur-342037, India

ABSTRACT

The primary focus of the method, developed here is on classifying soil types from satellite images with deep convolutional neural network. As we know there exist different types of soil around the land regions of the earth. But the physical and chemical properties of soil varies not only with change in regions, but also with time due to the factors like, deforestation, addition chemical waste etc. Here in this work we aim to develop a method to classify different types of soil based on their physical and chemical properties by analysing satellite images. Here four primary soil types, namely, alluvial, black, desert and red, available in India has been considered for our experimentation. An image repository, with around 1000 images, labelled with their soil-types, is also created for this work. The images, used here are of high resolution, acquired with LANDSAT-8 satellite. We have performed our study with these large number of labelled images from each soil type, captured in different time. We developed a method of soil classification with deep convolutional neural network. The merits of this method is experimentally validated with suitable comparison and proved to be effective in identifying proper soil class accurately.

Index Terms— Satellite Image Processing, Convolutional Neural Networks(CNN), Deep Learning, Image Segmentation, LANDSAT-8, Soil Classification

1. INTRODUCTION

Soil exists throughout the world in different varieties. Different types of soil have different behaviour and physical properties. These diverse behaviour and physical properties of the soil are suited differently to different types of agricultural requirements [5]. Besides, the properties of the soil in a particular geographic location may not always remain the same due to the factors like deforestation or addition of chemical waste. Therefore, there exists variation in soil properties spatially (with change in location) as well as temporally (with change in time).

The traditional methods for soil type classification involving soil surveys are quite expensive and time taking as it involves the challenges like large number of observations, on-field surveys and chemical analysis, and inconvenience in reaching some of the locations. Therefore those neither could be conducted frequently/ repetitively. However, the technological advancement in the fields of computer and information technology bring in totally new methods and new tools for carrying out such tasks computationally. The rapid growth in the fields of Remote sensing techniques and Geographical Information System (GIS) have opened new ways for estimating soil types and its spatial distribution with reasonable costs and accuracy [10].

In this paper, we propose a novel idea to classify the different soil types with deep convolutional neural network (CNN) and therefore be able to predict soil behaviour and its physical properties from the satellite image itself. We have developed a deep network for soil-type classification with convolutional neural network. Besides, we have created a repository by storing a large number of images acquired from LANDSAT-8 satellite according to their soil type. Around 1000 images are stored there from four soil classes. We used these images for training, validation and testing of our method. All the labelled images and the trained network can be found in the link [11].

The rest of the article is organised as follows. In Sec. 2 we discussed about some relevant work that involve satellite image processing. Sec. 2 describes our data collection and repository creation. The deep CNN architecture, used in this work is explained in Sec. 3. All the experimental results, along with suitable comparative studies are summarized in Sec. 4. The overall conclusion of the work is drawn in Sec. 5.

2. RELATED WORK

There exist a number of work on satellite image classification with conventional clustering and classification method [10]. Kumar *et al*. [1] first proposed a method for classifying soil type with satellite images with k-means clustering and k-nearest neighbour classifier. Phyo *et al.* [3] focused in classifying land cover regions from the images by integrating different features like RGB and L*a*b for the clustering and classification. Radhika and Varadarajan [4] developed a solution of satellite image classification with images of different resolution. Hamada *et al.*, [2] have performed pixel based clustering using the k-means clustering algorithm for corps classification from satellite images. Recently a few methods have been developed with deep neural networks for satellite image classification. Song *et al* [6] proposed a method for land cover classification with one dimensional CNN. Poliyapram *et al.* [8] developed a method of water, ice and land classification with deep learning. Tun *et al.* [9] proposed a method for different land cover region classification with CNN. Here in this article we have developed a method of soil type classification based on their chemical and physical properties with CNN. We have also created a data-base repository with proper soil-type labelling for that purpose as well. To our knowledge there does not exist any such data repository or trained deep networks for soil-type classification. In the next section we have described the creation of data repository in brief.

3. DATA REPOSITORY CREATION

The satellite images, that are used here in our work was downloaded from the USGS EarthExplorer repository [12], which gives a quick and intuitive way to download satellite imagery. The downloaded images were captured using the LANDSAT-8 satellite. The satellite uses two main sensors for capturing the images, namely Operational Land Imager(OLI) and Thermal Infrared Sensor(TIRS). OLI collects images using nine spectral bands in different wavelengths of visible, near-infrared, and shortwave light. It has sufficient resolution to distinguish features like urban centres, farms, forests and other land uses. While TIRS is mainly used for measuring water consumption. The metadata of all the downloaded images can be extracted from the name of the downloaded file itself. Fig. 1 shows the naming convention for retrieving the metadata. Each image represents an area of 185 km x 180 km on land. The resolution of the images are of the 7900x7900 pixels

.

LXSS_LLLL_PPPRRR_YYYYMMDD_yyyymmdd_CC_TX

Where:

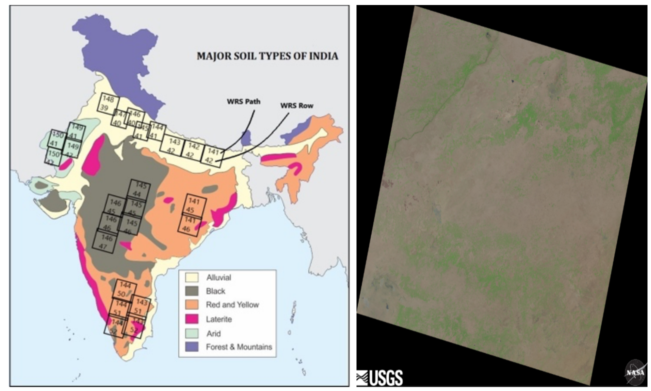
- L = Landsat
- X = Sensor (“C”=OLI/TIRS combined, “O”=OLI-only, “T”=TIRS-only, “E”=ETM+, “T”=“TM, “M”=MSS)
- SS = Satellite (”07”=Landsat 7, “08”=Landsat 8)
- LLL = Processing correction level (L1TP/L1GT/L1GS)
- PPP = WRS path
- RRR = WRS row
- YYYYMMDD = Acquisition year, month, day
- yyyymmdd - Processing year, month, day
- CC = Collection number (01, 02, …)
- TX = Collection category (“RT”=Real-Time, “T1”=Tier 1, “T2”=Tier 2)

Example: LC08_L1GT_029030_20151209_20160131_01_RT

Means: Landsat 8; OLI/TIRS combined; processing correction level L1GT; path 029; row 030; acquired December 9, 2015; processed January 31, 2016; Collection 1; Real-Time

Fig. 1. Naming Convention

Since the focus of our work was on soil-type classification, we collected and collated those images which are from the regions that belong to any one of the four major soil classes under consideration, namely, alluvial, black, desert and red. The approximate locations of the blocks, from which we collected the images are shown in Fig. 2(a). The black rectangles represent the blocks and the colour variations on the map reflects different types of soils present in around India. We have considered the images in visible range, that is in the bands 4, 5 and 6 which are equivalent to blue (B), green (G) and red (R) of our visible spectrum. We have collected 250 images from each soil type, and from different blocks. Our entire repository consists of 1000 images in total, 250 from each soil type.



(a) (b)

Fig. 2: (a) Approximate Locations of The Blocks from Each Soil Type (b) A Sample Sattelite Image

The name of the image in Fig. 2(b) is : LC08_L1TP_149041_20200318_20200326_01_T1.
The complete dataset is made available at this link [13].

4. ARCHITECTURE OF DEEP CNN

The underlying reasons behind our preference for Convolution Neural Network [14] to solve this task are as follows.
1. To extract out the relevant features for this multi-class classification task, 2.To capture the spatial distribution in the soil regions of images for the identification of their chemical and physical properties, and 3.To have the relevant feature map on the trained network for soil-type classification in future.

The architecture of the CNN used here is shown in Fig. 3. We have reduced the size of the input images to 1000X1000 pixels here, which are also high resolution images as well. Three convolution+ReLU layers and three pooling layers are used here prior to the flattened and fully connected layers. Each convolution layer is followed by a pooling layer. Masks of size 8X8 are used in the convolution layers. Max pooling is used in each pooling layer.

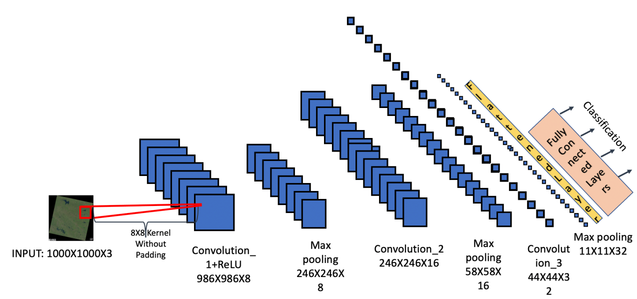


Fig. 3: Architecture of The CNN for Soil-Classification

The train-validation-test split was made in the ratio 70:10:20. We performed the experimentations by feeding different types of images in the input layer of the CNN. The variations that we have used in the CNN are listed below.

4.1. Raw Data as input

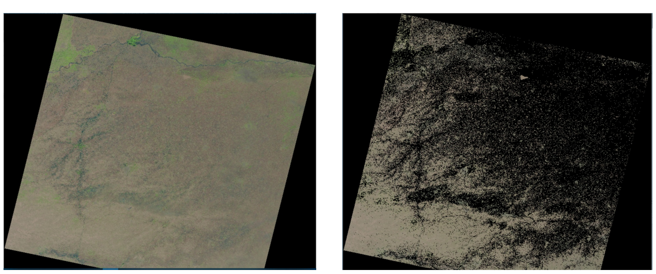
We fed raw images as input to the convolutional neural network constructed using different combinations of the convolutional layers, pooling layers and fully connected layers. We experimented on different combinations of layers, filter size in the layers and the number of filters used and obtained an average accuracy of 0.30 in 10 runs of the algorithm.

4.2. Absolute difference of temporal data as input

We trained the model with absolute difference of temporal images of the same block of soil. Absolute difference between temporal images of the same block was used as training and validation datasets. Median array was calculated from train images for each block and then the absolute difference of test images with median array were taken for building the test dataset. We experimented on different combinations of layers, filter size in the layers and the number of filters used and obtained an average accuracy of 0.34 in 10 runs of the algorithm.

4.3. Vegetation removed data as input

We performed image pre-processing steps on the raw images so as to remove vegetation cover before feeding the data to the neural network. Fig. 4(a) below is the original image and Fig. 4 (b) is how it looks after the segmentation is done on the image to remove vegetation. The experimental results of these variations are summarized in Sec. 5.1



(a) (b)

Fig. 4: (a) Original Image, (b) Image After Removing Vegetation

5. EXPERIMENTAL RESULTS

5.1 Variations In CNN Input Layer

As discussed in Secs. 4.1- 4.3, we trained the deep CNN with different types of input images and checked the accuracy. We have used DGX-2 for training the deep network. The network with vegetation removed images as input with different combinations of layers, filter size in the layers and the number of filters used and obtained an average accuracy of 0.67 in 5 runs of the algorithm with the maximum accuracy being as high as 0.80.

We also tried using vegetation removal step on one image from each block and used it as a mask for other images from the same block, but the accuracy decreased to 0.41 because of vegetation fluctuation with time.

The average accuracy, obtained after different experimentations are summarized in Fig. 5. As we can see from Fig. 5, the network with vegetation removed input layer is providing the highest accuracy. The rest of the experimental results, shown here were conducted with that network.

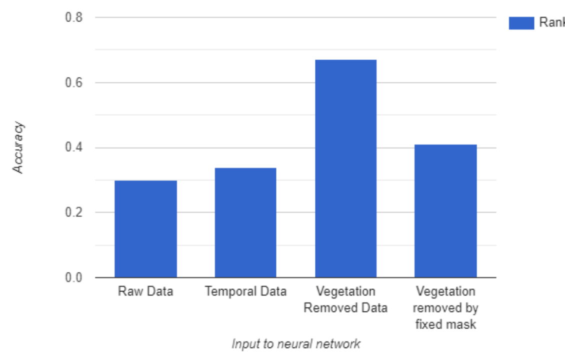


Fig. 5: Average Test Accuracy of CNN with Variation In Input Layer

5.2 Soil-Classification Accuracy of Vegetation Removed Deep CNN

We have trained and validated the proposed deep network with random separation of the data base into training-validation-testing in the ratio of 70:10:20, and run it for 15 epochs. We have repeated the experimentation five times, each time separating the data-base randomly. The average training (red colour) and validation (green colour) accuracies that we received with increasing number of epochs are shown graphically in Fig. 6(a), and the corresponding loss is shown in Fig. 6(b).

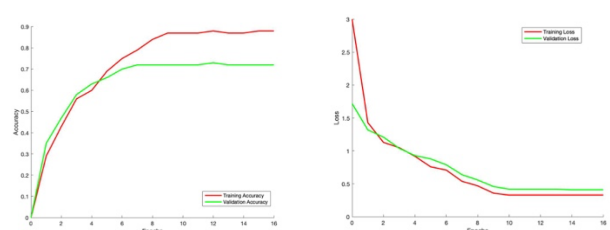


Fig. 6: Variations in (a) Accuracy and (b) Loss with Epochs In Red: Training and Green: Validation

We have tested our network in all the repetitions as well. As we know, we used 20% images from each class for testing, therefore, we had 50 images from each soil type, chosen randomly, used for testing. The average class-wise classification accuracy of testing is shown here in Fig. 7, in the form of a confusion matrix.

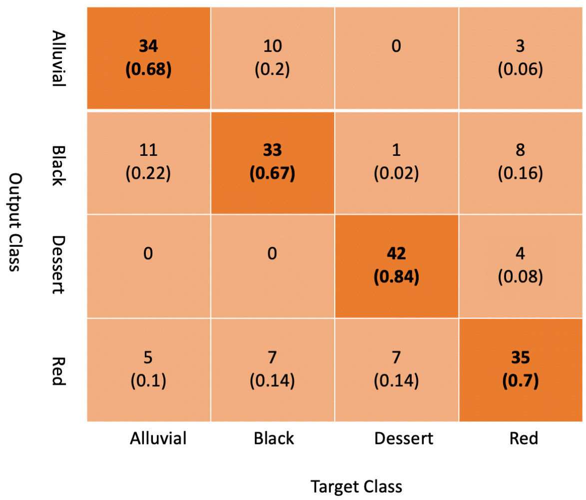


Fig. 7: Confusion Matrix of Class-wise Classification Accuracy in Testing

From the confusion matrix we can say that, the confidence on dessert soil class is the highest for this classifier, as there exists very few false positive and false negative points, whereas it is the lowest in case of black soil class, however the lowest value is as high as 0.67. The average classification accuracy of the network, during testing is 0.72, which is quite high.

For the sake of comparison we have implemented the conventional clustering and classification [1] on the same dataset considering colour and texture as the features. But there we get only the average accuracy only of 0.39, which is quite low compared to this deep CNN based method, proposed here.

6. CONCLUSION AND FUTURE SCOPES

The proposed method proves to be quite effective in classifying different types of soil from satellite images. This trained network can be used in future for soil class identification from satellite images. However we have not yet considered the regions where mixture of two or more types of soils are present. Besides, as we stated in Sec. I, soil properties could vary with time as well. Therefore are ample opportunities to explore this domain with the data-base that we have created. Besides, the data-base can also be enhanced with addition of more images and more soil classes.

6. REFERENCES

[1] N. S. Kumar, S. M. Anouncia and M. Prabu, *"Application of Satellite Remote Sensing to find Soil Fertilization by using Soil Colour", International Journal of Online Engineering (iJOE)*, vol. 9, no. 2, 2013

[2] M. A.Hamada, Y. Kanat and A. E. Abiche, *"Multi-Spectral Image Segmentation Based on the K-means Clustering", International Journal of Innovative Technology and Exploring Engineering (IJITEE),* Vol. 9, no. 2, 2019

[3] T. Z. Phyo, A. S. Khaing and H. M. Tun, *"Classification Of Cluster Area For satellite Image", International Journal of Scientific & Technology Research*, Vol. 4, 2015

[4] K. Radhika and S. Varadarajan, "Satellite image classification of different resolution images using cluster ensemble techniques," *IEEE International Conference on Algorithms, Methodology, Models and Applications in Emerging Technologies (ICAMMAET)*, 2017, pp. 1-6

[5] S. A. Siddiqui and N. Fatima, 'Indian Soils: Identification and Classification', *Earth Science India*, vol. 10, no. III, 2017, pp-1-14

[6] Y. Song *et al*., "Land Cover Classification for Satellite Images Through 1D CNN," *IEEE 10th Workshop on Hyperspectral Imaging and Signal Processing: Evolution in Remote Sensing (WHISPERS)*, 2019, pp. 1-5

[7] A. D. Vibhute, K. V. Kale, R. K. Dhumal and S. C. Mehrotra, "Soil type classification and mapping using hyperspectral remote sensing data," *IEEE International Conference on Man and Machine Interfacing (MAMI)*, 2015, pp. 1-4

[8] V. Poliyapram, N. Imamoglu and R. Nakamura, "Deep Learning Model for Water/Ice/Land Classification Using Large-Scale Medium Resolution Satellite Images," *IEEE International Geoscience and Remote Sensing Symposium (IGRASS)*, 2019, pp. 3884-3887

[9] N. L. Tun, A. Gavrilov and N. M. Tun, "Multi-Classification of Satellite Imagery Using Fully Convolutional Neural Network," *IEEE International Conference on Industrial Engineering, Applications and Manufacturing (ICIEAM)*, 2020, pp. 1-5

[10] S. Borra, R. M. Thanki, N. Dey, '*Satellite Image Analysis: Clustering and Classification',* Springer, Singapore, ISBN: 978-981-13-6424-2, 2019

[11] https://github.com/abhinaviitj/Soil-Type-Classification-From-Satellite-Images

[12] https://earthexplorer.usgs.gov/

[13]https://drive.google.com/drive/folders/1GCPKgiCfuEr9Fsffbou9o7bSpE9KI0Js

[14] R. C. Gonzalez, "Deep Convolutional Neural Networks [Lecture Notes]," in *IEEE Signal Processing Magazine*, vol. 35, no. 6, pp. 79-87, 2018