NDVI Versus CNN Features in Deep Learning for Land Cover Classification of Aerial Images

Anushree Ramanath, Saipreethi Muthusrinivasan, Yiqun Xie, Shashi Shekhar, Bharathkumar Ramachandra*
{raman074, muthu018, xiexx347, shekhar}@umn.edu, *bramach2@ncsu.edu
*University of Minnesota Twin Cities, *North Carolina State University*

***Abstract*—Agriculture plays a strategic role in the economic development of a country. Appropriate classification of land cover images is vital for planning the right agricultural practices and maintaining sustainable environment. This paper provides methods and analysis for land cover classification of remote sensing images. Satellite images form the input while mapping of every image to a distinct class is obtained as output. The objective is to compare the hand-crafted features based on Normalized Difference Vegetation Index (NDVI) and feature learning from Convolutional Neural Networks (CNN). The rationale of this work is to take advantage of techniques that are illumination invariant. NDVI versus CNN features have been compared on a linear Support Vector Machine (SVM). However, no comparative study has been carried out related to DVI based features and CNN based features on a deep learning classifier. This paper compares the performance of different classifiers and evaluates them based on test accuracy.**

Keywords—Normalized Difference Vegetation Index (NDVI), Multi Layer Perceptron (MLP), Convolutional Neural Network (CNN), Feature Extraction, Remote Sensing, Land scene classification, Aerial Images.

I. INTRODUCTION

Land cover classification of remote sensing images has a wide variety of application domains. It is mainly used in precision agriculture for soil, crop and pest management, land use planning, and water quality modeling. Changes in land cover and land use affect global systems (for example, atmosphere, climate, and sea level) or occur in a localized fashion in enough places to have a significant effect (Meyer and Turner, 1992). Hence, integrating data science with image analysis is crucial to understanding changes on a broad scale for the sake of creating better global environments in future.

Geospatial analysis is essential for a wide range of industry applications. Geospatial images and its classification form the key basis for these analysis'. Several Machine Learning and Deep Learning methods have tried various approaches in the past for classifying the geospatial images which include Decision Trees, Random Forests, Support Vector machines and so on. In this project, we intend to compare the performance of handcrafted features versus features extracted from Convolutional Neural Network (CNN). The features obtained from the two different techniques are fed into a Multi-Layer Perceptron (MLP) to learn the classification rule. Existing works compare handcrafted features such as Gray-Level Co-Occurrence Matrix (GLCM) and CNN features on a deep learning classifier. Others have demonstrated the effectiveness of techniques such as Normalized Difference Vegetation Index (NDVI) over CNN features on simple machine learning models such as SVMs, decision trees, random forests. Our study is an extension of this comparison where we look at NDVI versus CNN features on MLP for land scene classification. We intend to use the dataset SAT-6, which is based on sampling image tiles from the much larger National Agriculture Imagery Program (NAIP) dataset. A sample input of satellite images is as shown in Fig. 1. Validation of results is based on experimentation and comparison of the performance of different models is based on their test accuracy.

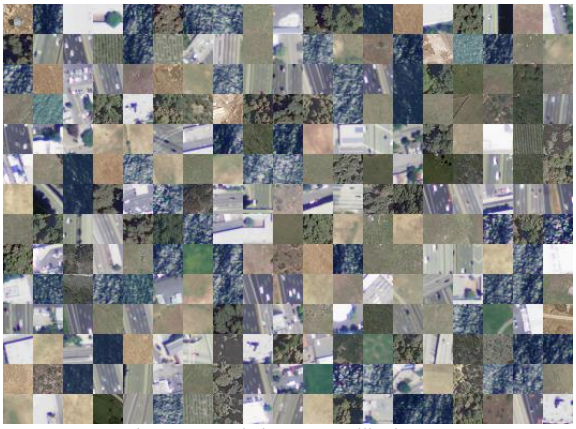

Fig. 1. Sample input - Satellite images

The aim of our study is to compare the performance of different classifiers for the dataset SAT-6, which comprises of six classes of land cover. An image consists of four bands - Red, Blue, Green and Near Infrared (NIR). The classifiers have the same underlying deep architecture - which is basically a Multi-Layer Perceptron. The classifiers will differ in the features of the input. First method will involve hand crafted features from the NDVI technique. The second method will have features extracted by CNN. The classifiers will be evaluated based on test accuracy.

The paper is organized as follows: In Section II, related work is presented. Section III describes the proposed approach. Validation methodology or experimentation details are presented in Section IV. This includes dataset, experimental design and CNN architecture. Results are presented in Section V. Finally, the concluding remarks are in Section VI.

II. Related Work

Both Machine Learning and Deep Learning methods have approached classification of geospatial images. [1] discusses several techniques for Multi-spectral broad-band vegetation indices available for use in precision agriculture from Remote Sensing images. NDVI, GNDVI are some of the manual feature extraction techniques popular in analysis of high resolution land cover images. In [2], it has been shown that feature extracted from pretrained CNNs performed the best when fed into linear SVMs. In [6], the paper compares different machine learning methods such as Decision Trees, Random Forests, Support Vector machines where the features are both object based, and individual-pixel based. While it is conventional to use raw pixels, sometimes it is more helpful to borrow techniques from Computer Vision to pre-process images or extract handcrafted features. In [4], the authors compare the results on OrthoPhotos for land cover classification by three different feature extraction methods in deep learning. In the first method, the features are nothing but the RGB channels of raw pixels along with the DSM (Digital Soil Model). The second method uses GLCM (Gray Level Co-occurrence Matrix) features. Lastly, the third method uses the features learned by a CNN (Convolutional Neural Network). These features were fed into an MLP (Multi Layer Perceptron) for supervised learning and CNN outperformed the other methods. Other feature learning methods include Bag of Visual Words (BoVW) to get textual context of an image for classification. Convolutional Neural Networks have been wildly successful in this area and papers [4], [5], [7], [8] highlight their effectiveness in various settings (supervised learning, layer-wise unsupervised pre-training and so on). There exist a few pre-trained CNN's such as ImageNet, AlexNet, RESNET which are known to be highly successful in certain image classification datasets that are comprised of certain classes of everyday objects. In [7], an interesting study was done to see if such deep features from pre-trained networks generalize well from everyday objects to aerial scenes domains. While CNN's performed well on the aerial images, they were still outperformed by low-level descriptors such as color and texture. RNN (Recurrent Neural Network) have also been applied for hyperspectral image classification [5].

While the above methods are examples of supervised learning, there exist unsupervised learning methods which have also been applied for remote sensing images. Autoencoder models learn a latent representation of input via a non-linear mapping [5]. PCA (Principal Component Analysis), SVD (Singular Value Decomposition) are classic dimensionality reduction techniques which also fall under unsupervised learning methods. In [8], the paper proposes a combination of greedy layer-wise unsupervised pre-training coupled with the Enforcing Population and Lifetime Sparsity (EPLS) algorithm for unsupervised learning of sparse features and shows the applicability and potential of the method to extract deep sparse feature representations of remote sensing images (sparse unsupervised deep convolutional networks). There also exist semi-supervised learning methods such as ss-kCCA, which have been applied on remote sensing images [9].

Finally, there are kinds of neural networks whose connections within the hidden/input/output layers are non-deterministic. Some examples are Deep Belief Networks and Restricted Boltzmann Machines [5]. They have not been as widely applied as the previously mentioned techniques in Remote Sensing context.

III. Proposed Approach

While NDVI vs CNN features have been compared, they have been done so on a linear SVM [2]. Our study is different in two ways. Firstly, we extend upon the CNN architecture constructed by Saikat et al [3] by introducing batch normalization layers for every convolution layer. Secondly, we use a Multi-Layer Perceptron, a deep learning network to compare the features while the previous study was done using a linear SVM. We chose to compare these two feature extraction techniques because, CNNs have demonstrated their effectiveness in extracting features automatically especially in images. We introduce a batch normalization layer after applying convolution. We chose NDVI since it is suited for identifying vegetation cover and its illumination invariance property. Our intent is to study the effect of normalization in the following techniques - learned from conventional CNNs, manually extracted by NDVI on an aerial images' dataset for classification. Multi- Layer Perceptron will serve as the underlying deep neural network to which the above-mentioned features would be fed. Our approaches are highlighted in the taxonomy diagram presented as Fig. 2 and 3.

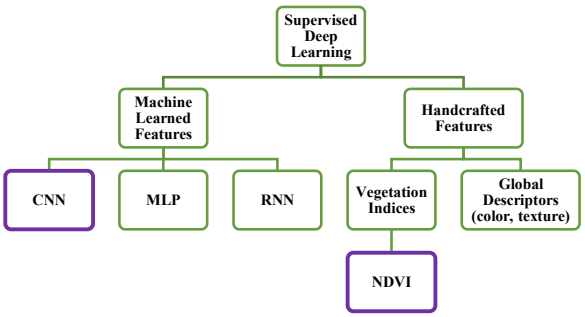

Fig. 2. Related work and novelty

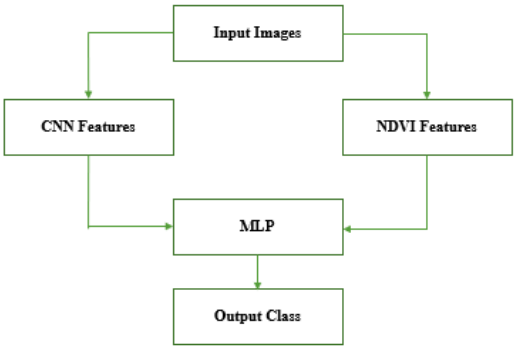

Fig. 3. Approach

IV. Validation Methodology: Experiments

A. Dataset

We used the dataset SAT-6, which was developed by Saikat et al [3] wherein the images are sampled from the much larger National Agriculture Imagery Program (NAIP) dataset [10]. The

entire NAIP dataset is ~65 terabytes spanning the whole of the Continental United States (CONUS). Saikat et al used the uncompressed digital Ortho quarter quad tiles (DOQQs) which are GeoTIFF images and the area corresponds to the United States Geological Survey (USGS) topographic quadrangles. To maintain the high variance inherent in the entire NAIP dataset, they sampled image patches from a multitude of scenes (a total of 1500 image tiles) covering different landscapes from California. SAT-6 consists of a total of 405,000 image patches each of size 28×28 and covering 6 land cover classes - barren land, trees, grassland, roads, buildings, and water bodies. The images consist of 4 bands – red, green, blue and Near Infrared (NIR).

B. *Experimental Design*

For the vanilla CNN model, we used the same experiment set up as Saikat et al, with a small change where we introduced a batch normalization layer for all the CNN layers in between the convolutions and the activations. Saikat et al used the model as follows: the first convolutional layer comprised of 6 feature maps, followed by a subsampling layer of kernel size 3*3 with average pooling, followed by convolutional layer with 12 feature maps, followed by subsampling layer of kernel size 5*5 with max pooling, finally collected to the output layer. The pooling windows were overlapping with a stride size of 2 pixels. The last subsampling layer is connected to a fully-connected layer with 64 neurons. The output of the fully-connected layer is fed into a 6-way softmax function that generates a probability distribution over the six class labels of SAT-6.

C. *CNN Architecture*

CNN architecture used for experimentation is as shown in Fig. 4.

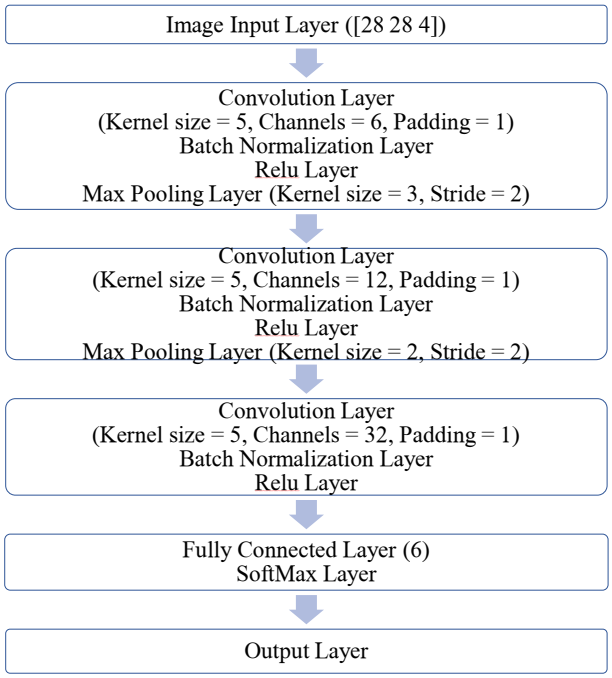

Fig. 4. CNN Architecture

For the Multi Layer Perceptron, we experimented with several architectures. The activation function used within the hidden layers is the hyperbolic tangent sigmoid function and for the output layer, it is the softmax function. The training function is the scaled conjugate gradient method, which was chosen for its computational efficiency in terms of speed.

V. Results

We use the work from Saikat et al to serve as baseline for comparing our CNN model. The baseline model achieved an overall test accuracy of 79% on the SAT-6 dataset.

For the MLP architecture of 1 layer deep 10 neurons wide network, we observed the highest performance. The results observed are described as follows:

For the hand-crafted feature selection method, we achieved an overall test accuracy of 83.25%. We observe much better results using our CNN network (Saikat + batch normalization). We achieved an overall test accuracy of 98.26%. The below confusion matrix (Fig. 5) provides a comprehensive view on the test data results. Interestingly, we see that the two feature extraction methods yield similar performance for the output classes except 4 and 5 (grassland and road respectively) where CNN performs significantly better. Table 1 presents the comparison of performance and Fig. 5 represents the confusion matrices for NDVI features with MLP and CNN features with MLP. In this case, MLP Architecture with 10 Neurons and 1 Layer is considered.

Baselines from Saikat et al -

- CNN: 79%
- Vegetation Index based feature + Deep Belief Network: 93.916%

TABLE I. Comparison of Performance

MLP Architecture	***NDVI Features - Overall Test Accuracy***	***CNN Features - Overall Test Accuracy***
1 Layer, 10 neurons	83.25%	**98.26%**
1 Layer, 100 neurons	84.8%	98%
1 Layer, 1000 neurons	72.6%	97.9%
2 Layers, 10 neurons per layer	**85.3%**	98%
10 Layers, 10 neurons per layer	84.8%	97.9%
50 Layers, 10 neurons per layer	62.9%	37.1%

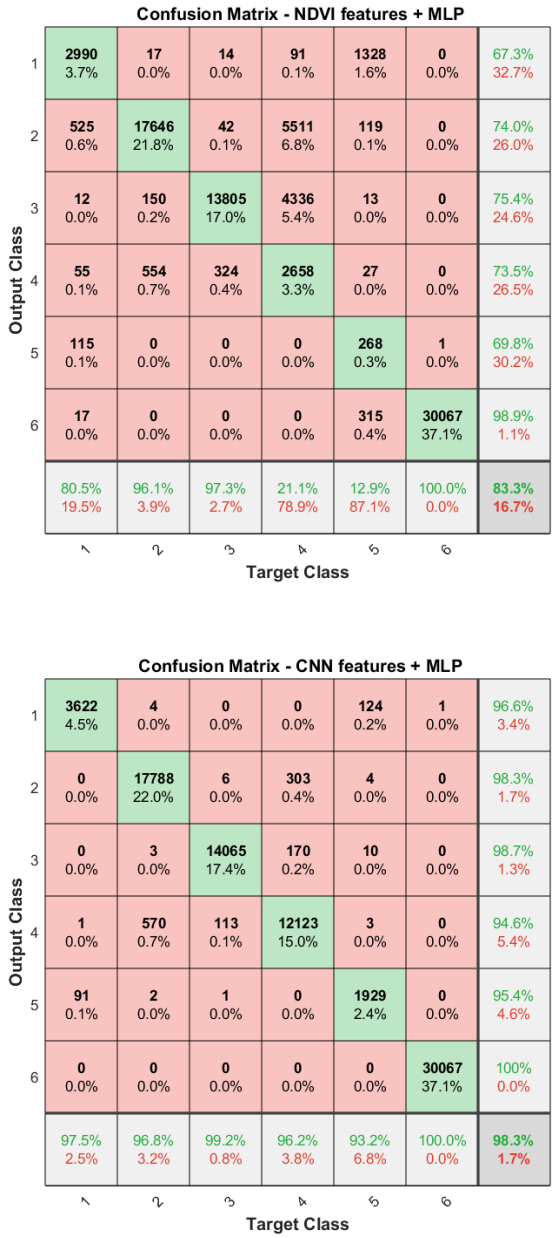

Fig. 5. Confusion Matrices – MLP Architecture (10 Neurons, 1 Layer)

We also looked at the generalization performance of both these techniques on classification of images taken in the dark. We used MATLAB's brighten function to darken the test images in the RGB channels to simulate pictures taken in the dark. We observed similar performance by both these feature extraction methods on the same test data with a much lower performance than what was observed for daylight images. Perhaps it could be that the method we used (MATLAB's brighten function only takes an RGB colormap as the input, and this ignores the NIR band) to simulate darkness was not appropriate.

VI. Conclusion and Future Work

In this paper, we have compared two different feature extraction methods - CNN and NDVI to see how automatic feature learning fares with respect to a manual feature extraction technique for image classification using a deep neural network, which is a Multi-Layer Perceptron. CNN features with batch normalization resulted in the best performance. We also see that NDVI is better than CNN without batch normalization. To this end, we have shown evidence to support that there is benefit in applying normalization for classifying remote sensing images, since they tend to vary in brightness as they span wide regions of space. Illumination invariance turns out to be the inevitable consequence of applying normalization. In future work, we think it would be beneficial to study the effectiveness of this technique to datasets entirely comprised of dark images.

References

[1] David J. Mulla, Twenty five years of remote sensing in precision agriculture: Key advances and remaining knowledge gaps, Biosystems Engineering, Volume 114, Issue 4, 2013, Pages 358-371, ISSN 1537-5110, https://doi.org/10.1016/j.biosystemseng.2012.08.009.(http://www.sciencedirect.com/science/article/pii/S153751 1012001419)

[2] Keiller Nogueira, Otvio A.B. Penatti, and Jefersson A. dos Santos. 2017. Towards better exploiting convolutional neural networks for remote sensing scene classification. Pattern Recogn. 61, C (January 2017), 539-556. DOI: https://doi.org/10.1016/j.patcog.2016.07.001

[3] Saikat Basu and (2015). DeepSat - A Learning framework for Satellite Imagery. CoRR, abs/1509.03602.

[4] J. R. Bergado, C. Persello and C. Gevaert, "A deep learning approach to the classification of sub-decimetre resolution aerial images," 2016 IEEE International Geoscience and Remote Sensing Symposium (IGARSS), Beijing, 2016, pp. 1516-1519. doi: 10.1109/IGARSS.2016.7729387R.

[5] Xiao Xiang Zhu and (2017). Deep learning in remote sensing: a review. CoRR, abs/1710.03959.

[6] Dennis C. Duro, Steven E. Franklin, Monique G. Dubé, A comparison of pixel-based and object-based image analysis with selected machine learning algorithms for the classification of agricultural landscapes using SPOT-5 HRG imagery, Remote Sensing of Environment, Volume 118, 2012, Pages 259-272, ISSN 0034-4257, https://doi.org/10.1016/j.rse.2011.11.020.

[7] O. A. B. Penatti, K. Nogueira and J. A. dos Santos, "Do deep features generalize from everyday objects to remote sensing and aerial scenes domains?," 2015 IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), Boston, MA, 2015, pp. 44-51. doi: 10.1109/CVPRW.2015.7301382

[8] A. Romero, C. Gatta and G. Camps-Valls, "Unsupervised Deep Feature Extraction for Remote Sensing Image Classification," in IEEE Transactions on Geoscience and Remote Sensing, vol. 54, no. 3, pp. 1349-1362, March 2016. doi: 10.1109/TGRS.2015.2478379

[9] J. Arenas-Garcia, K. B. Petersen, G. Camps-Valls and L. K. Hansen, "Kernel Multivariate Analysis Framework for Supervised Subspace Learning: A Tutorial on Linear and Kernel Multivariate Methods," in IEEE Signal Processing Magazine, vol. 30, no. 4, pp. 16-29, July 2013. doi: 10.1109/MSP.2013.2250591

[10] National Agriculture Imagery Program (NAIP) Information sheet: https://www.fsa.usda.gov/Internet/FSA_File/naip_2009_info_final.pdf