Land Cover Classification from Time Series Satellite Images

Mahesh Keswani
INFT Department
VESIT
Mumbai, India
2017.mahesh.keswani@ves.ac.in

Sanket Mahale
INFT Department
VESIT
Mumbai, India
2017.sanket.mahale@ves.ac.in

Rahul Kanwal
INFT Department
VESIT
Mumbai, Maharashtra
2017.rahul.kanwal@ves.ac.in

Dr. Shalu Chopra
INFT Department
VESIT
Mumbai, India
shalu.chopra@ves.ac.in

***Abstract*—The earth surface is continuously observed by satellites which are continuously orbiting around earth. This results in the generation of large satellite image data sets. This data is of increasing spatial resolution and temporal density. Land cover classification is an important application of satellite imagery. The results obtained from land cover classification is used to observe land use and changes such as resource planning, to detect deforestation, desertification and water scarcity. Land cover classification also helps in identifying the changes in the land cover. Most satellites along with high-end sensors provides pixel wise data. Due to high resolution images, the number of pixels which are generated are in millions even for the study of a small area. Therefore there is a need to balance runtime and accuracy for this task. In this paper, we did a comparative study of the models comparing Multi CNN model with traditional models to classify land cover from multi-spectral temporal Landsat-8 satellite data.**

***Index Terms*—Land Cover Classification, Landsat-8, Multi-Modal CNN, Machine Learning, Satellite Image.**

I. Introduction

Landsat 8[2] is a American satellite used for monitoring Earth and was launched in 2013 by NASA. It has a lot of sensors and equipment on board which can be used to capture data which was previously unaccessible by its previous generation. It Captures around 750 scenes a day. These images are available for download on their website. The sensors on the satellite captures 7 bands a combination of this band values can help us in detecting the type of land surface in the images.Taking this images over a certain period of time helps us in detecting the changes which have occured on the surface.

The images available are in .CSV format which is converted in a 3D Array and each cell of this array represents a spectral value along with few derived attributes of each pixel of 23 time steps, which describes the type of land cover associated with it. This land cover is the combination of various band values which are captured by the sensors in the satellite.

This pixel wise classification and using the images which are in time series format can help us exploiting the new patterns and behaviour per pixel and which in turn helps in prediction of land cover of pixel more accurately. These patterns are visible when we compare the data obtained from the images over a period of time. The model should be able to handle high volume of data as the total number of pixels can range over in millions. The input can also include data over a long period of time or classification is to be done over a large area which adds more pixels to be classified .

These Satellite Image Time Series (SITS) are a consequential source of information for scene (i.e. geographic area) analysis. A possible but verdant utilization of these images would consist in culling two images from the series and studying their differences and the evolution they reveal. However, vicissitudes in a scene might spread over a long duration (urbanization, for instance, lasts for several years and building sites do not have the same start time and culminate time) or they might cycle (such as crop rotation). Consequently, the number of possible coalescence is intractable and cannot be truncated to the analysis of two images

II. OBJECTIVES

The Following are the main objectives :

- Accurate Identification and Classification of each pixel value in the dataset.
- Identification of Model which gives satisfactory accuracy.
- Reduction of time required for classification thus balancing runtime and accuracy.

III. LITERATURE SURVEY

Multiple papers were studied and their findings are summarised in this section. This section includes papers studied before and during the development of the project. The papers helped in gaining insight into existing solutions, possible ways to optimize algorithms and facilitate the selection of algorithms based on their performance.
In [5] Efficient Satellite Image Time Series Analysis Under Time Warping, the authors F. Petitjean and J. Weber use the standard K-MEANS algorithm and then automatically label every cluster with regard to the most similar class. It also uses DTW to measure how the two optimally aligned series differ from each other. This was a long term approach for images over 20 yrs with lots of images hence analysis was done over a group of pixels and not individual pixels.
In [8], Minget et al. explores the classification approaches for faster and better prediction of land cover classification,

the approaches used were Random Forest and Genetic Algorithm. Random forest is robust approach based on statistical techniques for classification. They are also tested in the field of remote sensing. In this approach, selection of features is possible both experimentally as well as heuristically.

IV. Dataset Description

The dataset is from 23 Images from the Landsat 8 satellite which are obtained over as period of 365 days in 2014 above a Island to south of France called Reunion Island[1] this Dataset contains 2866 X 2633 pixels at 30 m spatial resolution provided at level 2A. We calculated 3 indices which are NDVI, NDWI and effulgence index - BI. This makes a total of 10 attributes which includes 7 original band values and 3 calculated indices.

A. Bands:

- Band1: Coastal
- Band2: Blue
- Band3: Green
- Band4: Red
- Band5: Near-Infrared
- Band6: SWIR1
- Band7: SWIR 2

B. Indices

- NDVI: The normalized difference vegetation index[10] is a simple graphical indicator of presence of green vegetation.

$$NDVI = \frac{NIR - Red}{NIR + Red} \tag{1}$$

- BI(Brightness Index): The Brightness Index Score[11] is a indication of fluorescence intensity of the soil.High BI indicates high salt content and soil humidity (1=dim, 5=brightest).

$$BI = \sqrt{\frac{Red^2}{2 * Green^2}} \tag{2}$$

- NDWI: Used to Monitor changes related to water content in water bodies[12].

$$NDWI = \frac{NIR–SWIR}{NIR + SWIR} \tag{3}$$

- Spectral features: [7] The spectral features (frequency based features), which are obtained by converting the time based signal into the frequency domain using the Fourier Transform, like: fundamental frequency, frequency components, spectral centroid, spectral flux, spectral density, spectral roll-off, etc.

C. Classes in Dataset

ClassId	Class Name	Instances
1	Urban Areas	16000
2	Other built-up surfaces	3236
3	Forests	16000
4	Sparse vegetation	16000
5	Rocks and bare soil	12942
6	Grasslands	5681
7	Sugarcane crops	7656
8	Other crops	1600
9	Water	2599

Fig. 1. Classes and their instances in Dataset

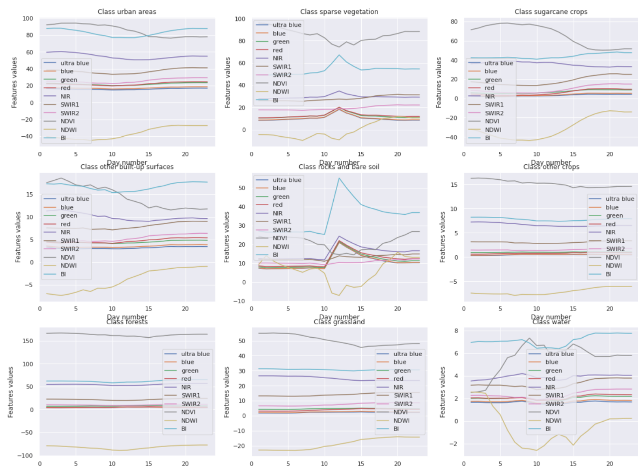

Fig. 2. Band Value variation according to Class

We performed exploratory data analysis on the data. One such result is shown in Figure 2. 9 graphs are for the 9 classes. In X-axis, the day number is there and in Y-axis, feature value is there. One graph shows the variation in the mean feature value of all the 10 features throughout 23 days. This shows how the different band values act according to the type of land cover. For example, water bodies have the highest NDWI value. Forest and Grasslands have high NDVI values.

As the dataset is band values from the satellite sensors it cannot be considerd as a image processing problem as we have access to pixel-wise band values and not the image as whole also due to timeseries part of the dataset we can view it as signal processing.It comes with pixel coordinates data which helps in localisation of time series space.

V. Machine Learning Approaches Applied

A. Random Forest

Random Forest[4] is a Machine Learning algorithm which is rooted on the concepts like Bagging, Bootstrapping and Decision Trees. Bagging is a technique of model aggregation, for classification taking the label which is most voted by k high variance base model. Every ith model is trained on bootstrapped dataset which betokens dataset is sampled from original dataset with supersession, additionally in random forest features are sampled which avails in minimizing the correlation between high variance and low bias models The Random Forest algorithm is insensitive to noise and due to bagging, it is quite resistant to overfitting leading to one of the best ensemble algorithms for classification purpose, furthermore it is faster in computation. We explore this algorithm for land cover classification on time series dataset. But Random Forest does have drawbacks as it does sampling with replacement and randomly splits features for each base estimator In this approach of per pixel classification, the dataset was rearranged daywise, but random forest is typical machine learning approach which do not takes into account any order of the features also every sample in time series data is considered as independent of its neighbor samples. Bootstrapping and random sampling will lead to bias in the out of bag error estimate of the forest if the time series has significant autocorrelation.

B. Support Vector Machine

Support Vector Machine[4] is a very powerful algorithm for relegation and it is a very good algorithm for plug and check on different dataset as it requires very little tuning. SVM finds the hyperplane which is proximate to the best estimate for distinguishing the types of data, in two dimensional planes, the hyperplane is line. SVM commences by plotting the data into D dimensional space, where D is the number of features, next it commences finding the optimal hyperplane which best segregates the data. SVM additionally works authentically well with non linear data as well, without any modification for linearly separable data. In non-linear data, the kernel is utilized, which is nothing but a measure of homogeneity between samples. In the context of land cover classification from time series dataset, we found that support vector machines even though it does not take into consideration the order of the samples, therefore it does not exploit the hidden patterns or behavior in the temporal domain. This implies that such methods are unable to reproduce or detect temporal trends in the data.

C. Deep learning in remote sensing

Deep learning is a sub field of Machine learning and has emerged as a huge success in the field of Machine learning in recent times[6]. It is based on the structure of the brain and has basic units as neurons and neurons collectively in the form of two hidden layers can approximate any function as proved by Universal Approximation Theorem. It has shown its capabilities in many areas like object tracking, translation into different languages, face recognition and many more complicated tasks.

The two important architectures of deep learning are Convolutional Neural network and Recurrent Neural Networks: Convolutional neural networks are a type of neural network which works on the idea of weight sharing and sparse connections. These two ideas are responsible for faster training as compared to traditional ANN and due to them along with some techniques, various popular architecture in the field of vision uses CNN for automatic filters finder. When multiple layers of convolutions are cascaded together, starting few layers will detect basic features like colors, edges, shapes and much more, but as we go deeper, layers start to capture patterns which are problem specific. One of the important parameters in CNN is shape of the kernel, which can decide how are we gonna perform using CNN, being a hyperparameter, we cannot take large shape, it can be computationally harder and smaller shapes can lead to underfitting for such a complex problem.

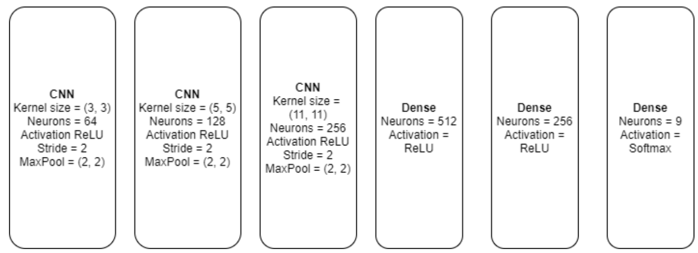

Fig. 3. Standard CNN Architecture

Figure 3 explains the architecture of a CNN model. It has several terms which are defined below.
ReLU: Rectified Linear Unit is non linear function which is defined as max(0, x).
Neurons: The total number of neurons in each layer.
Kernel: The kernel is basically a filter which is used to capture particular latent representation of input and it is learnt iteratively by optimizer.
Stride: It is basically number of steps the kernel moves along x and y direction during convolution.
Pooling: Pooling is one of the way for downsampling the feature maps leading to lesser but important number parameters of network.

From Figure 3 we can see that, for the land cover classification task on time series images, the convolutional neural network was able to outperform the traditional random forest and support vector machines, especially for the samples with the classes whose count was very small as compared to the other classes.

However, in CNN, while using the kernel of same size as hyperparameter, we are restricting ourselves to detect patterns in the time series images, therefore we can exploit it using by applying convolution with filters of different shapes on the same input and concatenating the feature maps along the channel dimension, which is what we propose in our solution.

VI. Proposed Multi Modal CNN1D

In Multi Modal one dimensional convolutional approach[9], we make use of four different convolutional networks, as shown in Figure 4.

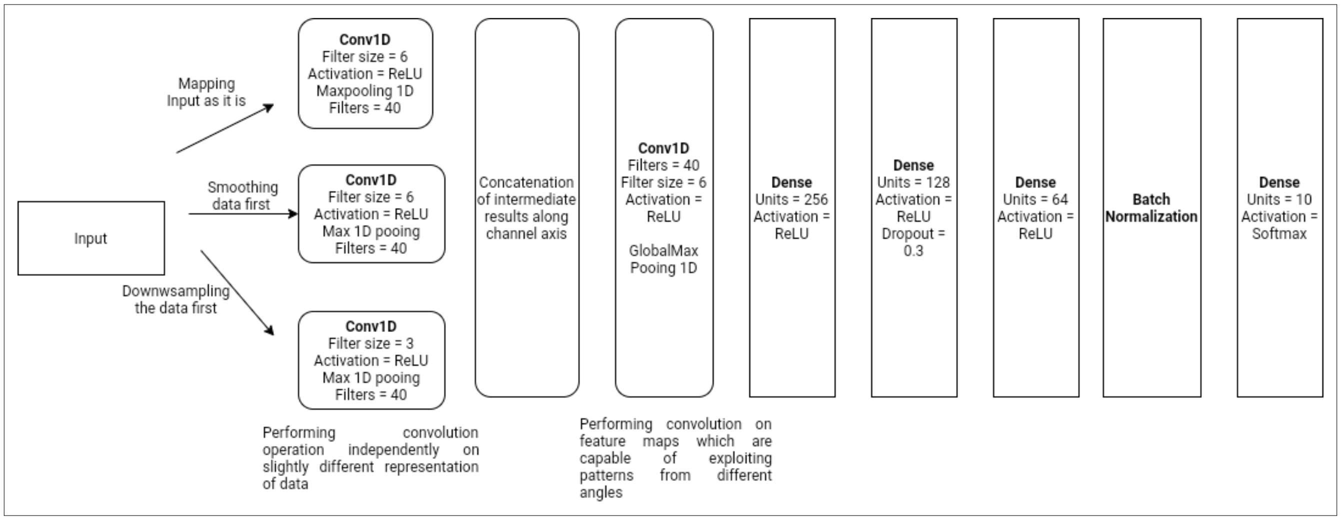

Fig. 4. Multi Modal CNN Architecture

For the first network, the input is directly passes as it is, same as mapping identity, while for the second network, instead of using the input as it is, we first smooth the data and then perform the local convolution and for the third network, we first down sample the data by taking the sample from input with sampling rate as 2.

For all these independent convolutions, we use 1-D convolution on each of these newly generated time series data. One of the important benefit of this approach is that the intermediate results which are captured from the local convolutions are quite robust to scale of time series data. Also due to the downsampling, the computational complexity is reduced to great extent as we are not in continuously increasing size of kernel, using the same filter size by which the number of parameters to be learnt is reduced and also leading to lower complexity which can resulting in better generalisation.

After all the convolutions are done on a different scale of original time series, we concatenate all the feature maps along the channel dimension and by using the great latent representation of the input, we perform convolution on them following the traditional artificial neural network for approximating the underlying function.

VII. Results

Figure 5 shows the results for every label wise accuracies for every model on which we performed the experiment on time series satellite images. The metric used for the accuracy is F1-score. We tested the models on the test data, which was for the two time steps.

By observing the results, the choice of classifier doesn't make much difference. All the classifiers gave a lower score for Other built-up surfaces and other crops. One reason for this can be to the fact that the instances of the other built-up and other crops class is the lowest.

MCNN model gave relatively higher scores than other models. This can be attributed to the fact that MCNN explores the data from different angles. So it doesn't depend solely on the frequency of the class. But we are observing that MCNN too gives a lower score if the frequency of the class is very low. In fact, standard CNN gave a higher F1 score on Other built up class than MCNN.

Land Cover	RF	SVM	CNN	MCNN
Urban	79.1431	81.0207	79.8717	90.5217
Other built-up	59.6265	62.2093	69.0541	68.6721
Forests	75.5033	78.8606	75.7162	89.2125
Sparse Vegetation	65.3264	69.3705	83.3747	91.0251
Barren	62.0129	71.3744	85.9080	94.2432
Grassland	81.6973	81.9032	79.3735	81.3507
Sugarcane	67.9366	71.5958	76.0662	92.2305
Other Crops	59.3600	65.9483	70.5668	68.0985
Water	70.13111	73.7243	81.7387	76.5021

Fig. 5. Classes wise F1 Score on RF, SVM, CNN and MCNN Models

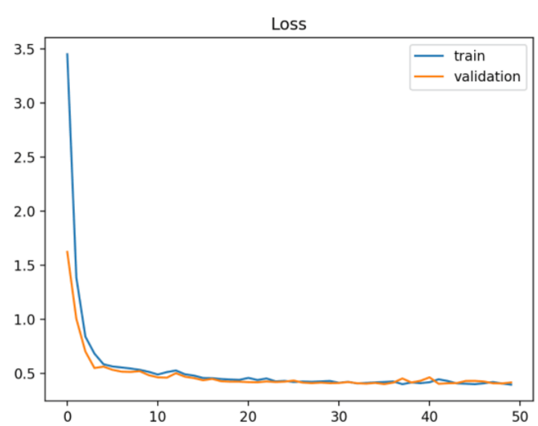

Fig. 6. Loss of MCNN

VIII. Conclusion

In this solution, we did a comparative study of Random Forest, SVM, CNN and Multimodal CNN model for the purpose of Land Cover Classification on SITS data. We used 23 Landsat-8 images of Reunion Island which had 9 classes. In our study, we found out that Multimodal CNN outperformed the other three models. Using MuliModal CNN Model and preprocessing the accuracy obtained was the highest. This Model and with the preprocessing techniques, it can be used in pattern recognition or other classification problems even for the classes which are imbalanced as compared to other classes. As the class results obtained can be used for calculating percentage change and change of forest or sparse vegetation can be classified as deforestation in a similar way in an increase in Urban class is Urbanisation and the percentage increase in rocks and bare soil is Desertification.

Acknowledgment

We would also like to thank Mr. Dino Ienco and UMR TETIS, France for explaining the problem and providing the dataset for training and testing.

References

[1] "TiSeLaC : Time Series Land Cover Classification Challenge - Dino Ienco," *Google Sites*. [Online]. Available: https://sites.google.com/site/dinoienco/tiselc. [Accessed: 18-Mar-2021].

[2] "Landsat Missions," *Landsat 8*. [Online]. Available: https://www.usgs.gov/core-science-systems/nli/landsat/landsat-8?qt-science_support_page_related_con=0qt-science_support_page_related_con. [Accessed: 18-Mar-2021].

[3] F. Petitjean, P. Gançarski, F. Masseglia, and G. Forestier, "Analysing Satellite Image Time Series by Means of Pattern Mining," *Intelligent Data Engineering and Automated Learning – IDEAL 2010*, pp. 45–52, Sep. 2010.

[4] P. Schafer, D. Pflugmacher, P. Hostert, and U. Leser, "Classifying landcover from satellite images using time series analytics," in *EDBT/ICDTWorkshops*, 2018, pp. 10–15

[5] F. Petitjean and J. Weber, "Efficient Satellite Image Time Series Analysis Under Time Warping," in *IEEE Geoscience and Remote Sensing Letters*, vol. 11, no. 6, pp. 1143-1147, June 2014, doi: 10.1109/LGRS.2013.2288358.

[6] B. Zhao, H. Lu, S. Chen, J. Liu and D. Wu, "Convolutional neural networks for time series classification," in *Journal of Systems Engineering and Electronics*, vol. 28, no. 1, pp. 162-169, Feb. 2017, doi: 10.21629/JSEE.2017.01.18.

[7] N. Hossain, "What are the Spectral and Temporal Features in Speech signal?," *ResearchGate*, 06-Mar-2015. [Online]. Available: https://www.researchgate.net/post/What-are-the-Spectral-and-Temporal-Features-in-Speech-signal. [Accessed: 18-Mar-2021].

[8] D. Ming, T. Zhou, M. Wang, and T. Tan, "Land cover classification using random forest with genetic algorithm-based parameter optimization," *Journal of Applied Remote Sensing*, vol. 10, no. 3, p. 035021, 2016.

[9] Z. Cui, W. Chen, Y. Chen, "Multi-Scale Convolutional Neural Networks for Time Series Classification", *ArXiv* abs/1603.06995 (2016)

[10] "Normalized difference vegetation index," *Wikipedia*, 12-Mar-2021. [Online]. Available: https://en.wikipedia.org/wiki/Normalized_difference_vegetation_index. [Accessed: 18-Mar-2021].

[11] *Brightness Index*. [Online]. Available: https://www.biolegend.com/en-us/brightness-index. [Accessed: 18-Mar-2021].

[12] "NDWI Normalized Difference Water Index," *Sentinel*. [Online]. Available: https://custom-scripts.sentinel-hub.com/custom-scripts/sentinel-2/ndwi/. [Accessed: 18-Mar-2021].