17. Primary Level of Land Use Land Cover Classification using AI/ML

Introduction

Land Use / Land Cover (LULC) describes the earth's characteristics and indicates how the land is utilized for various activities. Land cover change is a continuous process associated with urbanization, deforestation, drying of lakes, overutilization of land for agriculture, etc. It thus constitutes an essential basis for economic planning and resource management. Still, preparing LULC data in a spatial domain is time-consuming and laborious work involving a lot of human resources. Categorizing LULC of an area is a significant challenge while using Remote Sensing satellite data. Over the past decade, Machine Learning (ML) has seen an increasing uptrend and great interest due to its powerful ability to learn representations of data samples with several levels of abstraction on multiple processing layers. The categorization of land features using ML approaches due to its consistent output and less requirement of human intervention is the right approach and current trend in the geospatial domain.

ML Techniques for LULC Classification: Supervised algorithms like Support Vector Machine (SVM), Random Forest (RF), Maximum Likelihood classifier (ML), and deep Artificial Neural Network (ANN) are some common approaches to extracting thematic information from multispectral satellite images.

1. In 2001, Breimanproposed Random Forest (RF), an ensemble classification method consisting of many decision tree classification models with a random selection of training datasets and factors. The two most essential input criteria for this classifier are the size of the training dataset and the number of trees generated.
2. Maximum likelihood classifier (ML)is a supervised classification method that describes every band by a normal distribution. This supervised classification method is based on the Byes theorem.
3. Support Vector Machine (SVM)employs an optimization algorithm to locate the optimal class boundaries. SVM was initially designed as a binary linear classifier and further extended to deal with non-linear classification using a non-linear kernel function. In the training phase, SVM classifiers generate an ideal hyperplane that splits several classes with the less misclassified pixels from input datasets, ensuring that the width of the margin will be maximized.
4. The Artificial Neural Network (ANN) is the most widely applied machine-learning technique, which can be efficiently used in non-linear phenomena such as LULC changes with the ability to work on big data analysis. The ANN is a forward structure black-box model trained by a back propagation algorithm. The ANN has one input layer, at least one hidden layer, and one output layer. Each layer is formed by the weighted interconnection of non-linear processing units called neurons. The main function of the back propagation algorithm is to minimize the error between the actual network outputs and the outputs of training input/output pairs with the learning rate and updating the weights of the backward paths.

Methodology

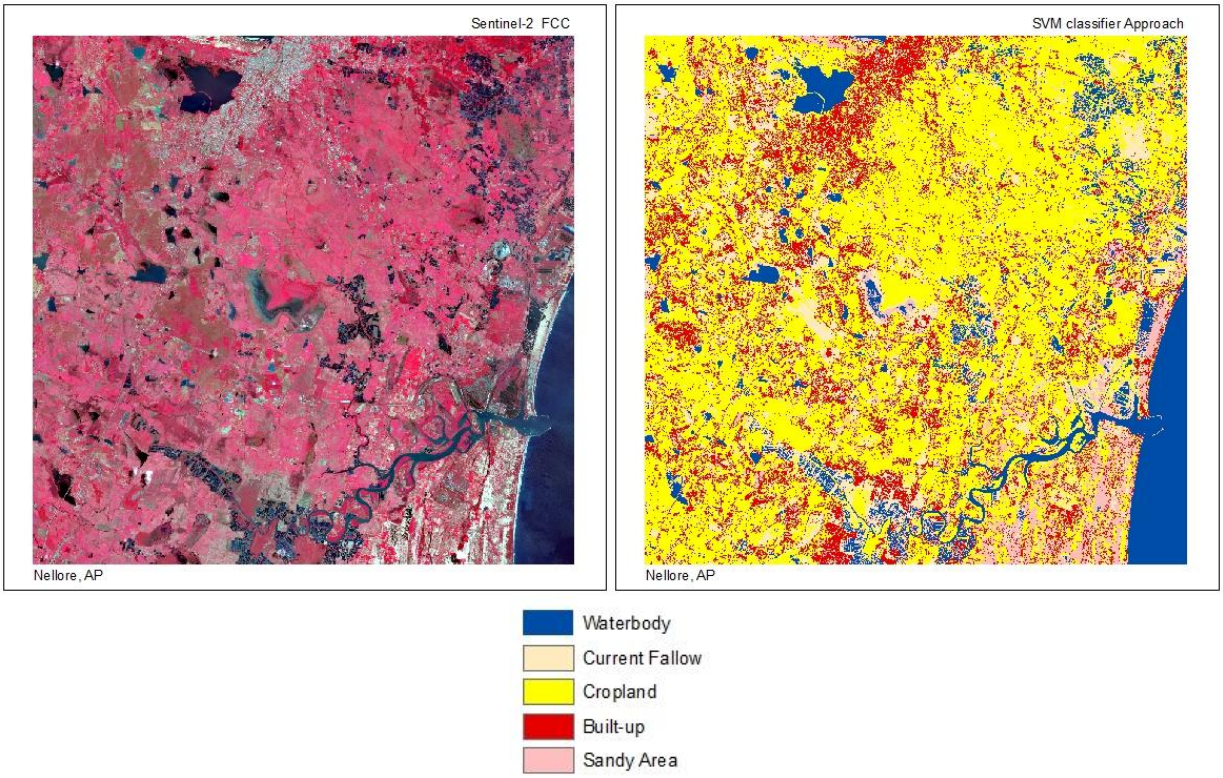
The following method is adopted for LULC classification using ML.

- Collect the remote sensing data and apply all necessary pre-processing steps.
- Collect training data with prior knowledge of place/location with ground truth points.
- The sample data is split into three sets: training, validation, and testing.
- Define a model with predefined necessary parameters.
- Train the classifier using the training data.
- Modify parameters according to validation accuracy.
- Classify an image or feature collection.
- Estimate classification error with test data.

Outcome

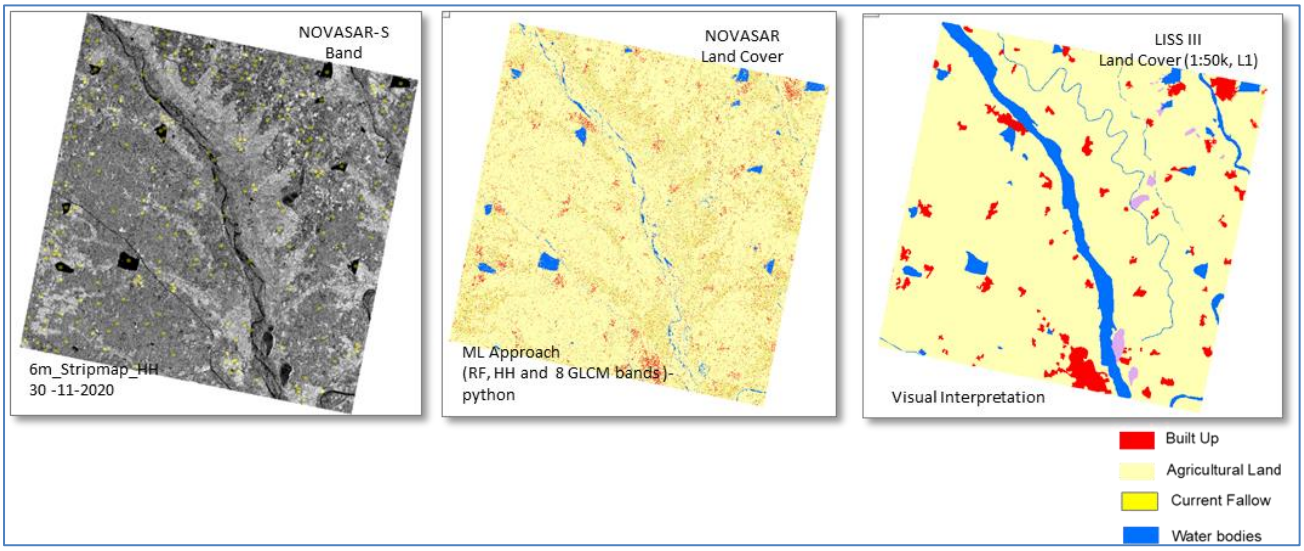
A classified image is obtained from the input satellite data using the model developed/defined. Two examples are shown here with different models and input data.

Case-1: Sentinel-2 optical FCC image with bands B8, B4, and B3 is used for the study site of Nellore and environs, Andhra Pradesh. ML technique of SVM is applied to the FCC to extract the primary Land cover classes. Results show that the SVM classifier provides satisfactory output for Sentinel-2 data.



Case-2: NOVASAR S-Band Stripmapdata with polarization HH is used for the study site at Gudur and environs. Level-2 data (sigma naught) is subjected to Land Cover

classification using machine learning techniques, and its comparison with other image classification techniques has been carried out. Preliminary result shows that NOVASAR-S-band data is amenable for primary Land Cover extraction. A customized model using Python is developed to create the model, tune its parameter, and classify it using texture-derived parameters and backscatter information.



Current Constraints/Challenges: Harmonisation between the area derived from two classification techniques is always challenging. Since the ML techniques classify using the pattern, reflectance, and texture of the remote sensing imagery, the area under each land-use class for the same region also varies in the different satellite data due to the atmospheric effects, illumination, and geometric variations. Training models require a considerable amount of pure training datasets for a better outcome. Parameter optimization of the ML models, techniques, and the accuracy differences in the algorithms used vary from user to user.