Name: Om Vaknalli

Roll No. 18376

Date: 30/03/21

Subject: EES – 336, Remote Sensing and GIS, Assignment 3

Aim: To classify a given dataset along the different types of land use classes using different Machine Learning techniques.

Procedure:-

1.) I downloaded the latest Landsat 8 images of Mumbai (my hometown) using the USGS Earth Explorer. The images that were downloaded were of bands 2, 3, 4, 5, 6 and 7.

Note: The Landsat images used were from the "LC08_L1TP_148047_20210314_20210314_01_RT" series.

2.) A mask was created which selected the Mumbai AOI appropriately.

3.) The mask created in the previous step was used for extracting the AOIs from each of the 6 band images.

4.) Each of the AOIs (of all bands) were then processed for atmospheric correction (DN to radiance by Gain & Bias method and radiance to reflectance by ToA correction)

5.) These corrected AOIs of different bands were then combined to create an FCC (with the symbology of an NCC).

6.) The FCC was then put through PCA for dimension reduction purposes.

7.) Now, the resultant image was processed through the Iso-Cluster Unsupervised Classification (ICUC) using 25 different classes. These classes were later reclassed based on manual observational comparison of the original bands, google earth images as well as the base map.

8.) The accuracy assessment points (AAP), frequency table and confusion matrix along with the corrected ground truth values were deduced for the ICUC outputs. Consequently, the user, producer and overall accuracies were also found along with their corresponding commission, omission and Kappa coefficients.

9.) The land-cover area and their discrepancies for each of the classes were calculated. The total area was also calculated.

10.) The result of step 6 was also put through supervised classification (SC) by means of selecting 100 training polygon datasets for each of the classes mentioned above. The steps 7, 8 and 9 were again repeated for this method.

11.) Maps of results of both the datasets were constructed and compared along with their accuracy parameters and land-cover differences.

Note: The above procedure is shown through extensive detail using the screenshots of the process. They are separately attached for verification purposes.

Observations & Remarks:-

1.) The date of the band images was carefully chosen to accommodate minimal amounts of cloud cover. This was done by using the 'Cloud Cover Filter' option on USGS Earth Explorer. The cloud cover in the raw data used is at most 1%.

2.) The process of Dark Object Subtraction was also attempted along with Atmospheric Correction on the above dataset by was unsuccessfully carried out due to logistical errors.

3.) Although an FCC was created using all the 6 downloaded bands, the symbology used only the visible spectra band images so that the visualization of the dataset can be done using a NCC analogy which helps in identifying map features in further procedures such as reclassing the ICUC and correcting ground truth values in SC.

4.) A mini exercise was also performed wherein the ICUC was done using varying no. of land-use classes (5, 10, 15, 20, 25 & 30 classes), and their results were compared. It was observed that selecting 5 & 10 classes yielded in extremely poor classification (due to complexity of the dataset), 15 & 20 classes gave moderately good results while 25 classes gave the optimal results. It was noticed that choosing 30 or more classes produced a range of grayscale class symbology which was tedious to analyze during the process of reclassification.

5.) It was observed that my choice of AOI had negligible landcover for agricultural and croplands. Hence, this class was neglected from the study. The other classes that were included were Waterbodies, Swamp & Vegetation, Open Areas and Urban Areas.

6.) There were some distinct land features that of the topography that both the classification methods failed to fully realize. For eg., the roads in the topography were sometimes identified as urban areas while other times as Open areas. A similar situation was observed with extremely dense patches of vegetation, which were frequently classified as open areas and vice versa. A judgement call had to be made while reclassing/correcting these points.

7.) The waterbody land-use class had the best accuracy through both the machine learning methods simply because the Mumbai AOI spanned about 40% of waterbodies (Arabian Sea, Powai Lake, Vihar Lake, Tulsi Lake and several other rivers and ponds). This made it easier to collect larger samples of training data in these regions for ICUC. The rest of the classes had similar order of errors. The complexity of the images was high, i.e., the sorting of the various land-use elements was very poor, which led to these results.

8.) In SC, the user accuracies of waterbodies, vegetation and open areas were greater than the producer accuracies of the same. This signifies that the algorithm identified lesser points in these land use classes than that which were actually present. On the other hand, the urban area showed a higher producer accuracy compared to the user accuracy, which meant that a higher percentage of points were excessively identified in this class. A similar logic can also apply to the ICUC wherein the waterbodies and urban areas show similar trends as above while the vegetation and open areas accuracy trends have reversed when compared with SC. The lesser the difference between the user and producer accuracies, the better is the classification of that specific class. Hence, when the table of accuracy values are compared for the methods of classification, the winner is quite obvious; it is the unsupervised classification.

9.) The above point can also be more directly noticed when comparing overall accuracy as well, where the accuracy of the former (SC) is greater than the accuracy of the latter (ICUC). This is fairly intuitive as we know that SC works better when the data being analyzed in heterogeneous and poorly sorted/hetero-clustered (seen with urban area accuracy comparison), while ICUC works better when the data is more homogeneous and well sorted (as can be seen in the comparison of waterbody accuracies). Considering that our data was more towards the heterogeneous side, it is only plausible to have better overall accuracy and Kappa coefficient of the SC analysis.

Results:-

1.) **Link to the google drive containing the screenshots and all supporting files of the procedure**

Note: The drive does not contain the original, RAW Landsat images.

2.) **Link to the excel file of Area and Accuracy analysis**