VOL. 17, NO. 11, JUNE 2022 ISSN 1819-6608
ARPN Journal of Engineering and Applied Sciences

www.arpnjournals.com

LAND USE LAND COVER CLASSIFICATION USING MULTI-SPECTRAL SENTINEL-2B SATELLITE IMAGE

N. Anusha[1], Sneha B.[2], Jyothi M.[2], Vihasitha G.[2] and Saisri N.[2]
[1]Department of Computer Science and Engineering, Vidya Jyothi Institute of Technology, Hyderabad, Telangana, India
[2]Department of Computer Science and Engineering, G. Narayanamma Institute of Technology and Science (For Women), Hyderabad, Telangana, India
E-Mail: anusha.nallapareddy@gmail.com

ABSTRACT

Land Use Land Cover (LULC) classification plays a key role in sustainable planning, management, and supervision programme at the local, regional, and national level. With a high level of precision, in this work the Principal Component Analysis (PCA) is utilized to suppress the noise and extract the characteristics from the Sentinel-2B satellite data. The resultant components of PCA are given as input to the K-means clustering algorithm. K-means clustering algorithm clusters the similar pixels into groups. This research performs LULC classification of a multispectral satellite image by utilizing Random forest classification method. The study area considered in this paper is the area covering in and around G. Narayanamma Institute of Technology and Science (GNITS), Shaikpet, located in Hyderabad, Telangana, India. Optical image obtained by Sentinel-2B satellite on February 27, 2021 is used in this study. To conduct a subjective evaluation of the clusters obtained by K-means clustering and to identify the LULC classes, the Google Earth Engine perspective of the study area is used. Classification of LULC performed by applying Random Forest model on the testing data resulted with an accuracy of 87.60%.

Keywords: classification, land use land cover, pre-processing, principal component analysis, random forest classifier.

1. INTRODUCTION

Different organizations use LULC maps extensively for various applications like urban planning, sustainable planning, etc. The process of categorizing land cover pixels into classes is known as LULC classification [1]. For monitoring changes on the surface of the earth, utilization of the remote sensing data is a very efficient and reliable source. Information about the LULC dynamics can be obtained by applying the image processing techniques on the remote sensing data [7]. LC refers to the physical characteristics of the surface in the distribution of manmade structures, natural resources like water, soil, rock and other features of the land, both naturally and artificially. LU refers to the purpose the land serves. It determines what sort of community, and environment can be used on a specific type of land. Remote sensing from space combined with Geographic Information Systems (GIS) has been for a long time regarded as a potent and effective technology for identifying LULC changes [4].

In general, LULC maps are created by classifying the images through remote sensing (satellite images or aerial photographs). These LULC maps are vital inputs for various developmental, environmental and resource planning applications, and regional as well as global scale process models [4].

1.1 Image Pre-Processing

After data collection, pre-processing is a crucial step that aids in extracting trustworthy information from satellite data. The most frequently used pre-processing methods in the literature are spatial image enhancement, atmospheric correction, geometric correction, radiometric correction, and topographic or terrain correction. In enhancing satellite images, pre-processing is an essential process and it assists in the further processing of satellite data. PCA and Linear Discriminant Analysis (LDA) enhance the satellite image through dimensionality reduction [10]. One of the best ways to improve the object features from multi-spectral satellite images is to use PCA and morphological techniques [5].

Prior to calculating the reflectance values at the pixel level, image processing is performed initially which includes absolute radiometric adjustment. Translation of digital values into radiance units is done by applying solar correction using Landsat metadata file settings. Data redundancy is removed by performing the PCA [12]. The quality of the image data is improved by the pre-preparing stage by decreasing or eliminating a variety of mistakes caused by both internal as well as external factors. On the image, histogram equalization can be applied for adjusting the contrast of the image [15].

The PCA is a set of approaches that takes high-dimensional data and converts it to lower- dimensional data without losing any of the original information. It is quite tough to interpret more than two variables and to reduce the collection of variables to new variables. This is something that PCA could perform. Principal components are a generic strategy that uses complex mathematical principles to minimize the number of linked variables to a smaller number of uncorrelated variables. It is a straightforward and reliable procedure known as the Karhunen-Loeve (KL) Transform and the Hotelling Transform [5]. PCA is responsible for dimension reduction [5, 10], compression, augmentation, and noise removal [15]. PCA reduces the number of connected variables to a smaller number of unrelated variables known as primary components. The first main component has a huge amount of variation, whereas the subsequent components have as little variation as feasible. PCA is a

statistical approach for revealing a set of variables' covariance structure [5].

The Sentinel Application Platform (SNAP) is used to perform pre-processing. It is developed for the European Space Agency (ESA) by CS in collaboration with CS-Romania, Brockmann Consult, INRA, UCL and Telespazio Vega Deutschland [4]. The open-source Quantum GIS (QGIS) software (Version QGIS 3.0.0) is used for performing the image processing chain. The semi-automatic classification plug-in is used to apply atmospheric correction based on the Dark Object Subtraction (DOS) technique [11].

The visual appearance of satellite data is affected by issues such as noisy or foggy data, geometric, topographic faults and dead pixels. So, pre-processing is an important step following data collecting, as it aids in the extraction of trustworthy information from satellite data [10]. Satellite image enhancement is the most important approach in the world of satellite image processing for boosting feature visibility. Digital image processing techniques are necessary to increase the quality of the objects in order to obtain better feature extraction of the object values [5].

The software tools that are used to process remote sensing images are: PolSARpro - to process the polarimetric radar images. The Next ESA SAR Toolbox (NEST) (new name: SNAP/ Sentinel-1 Toolbox) is a toolbox developed by the European Space Agency (ESA) for speckle filtering and range doppler terrain correction. ENVI 5.0 is used for co-registration of the FORMOSAT-2 image [7].

1.2 Image Processing

Image analysis enables to extract the information from data sets. Using a variety of approaches or a combination of techniques, different datasets can be processed separately namely Hyper spectral data processing (Hyperion and AVIRIS-NG) and multispectral data processing (Sentinel-2 and Landsat 8 OLI) [8].

Every pixel in the image dataset is assigned to a spectral class, which resembles a mathematical decision rule. When it comes to classifying pixels, there are two primary approaches: supervised and unsupervised [10, 15, 16]. A classifier is used in supervised classification, and each class requires a training sample. Statistical clustering methods are used in unsupervised classification to select spectral classes inherent in the data [15].

Cluster analysis is a popular method of non-supervised learning. The goal of cluster analysis is to identify hidden structure in the data and group the data of similar kind together as much as possible using similarity measurements [18].

ISODATA clustering, an unsupervised classification method is applied on multi-temporal [2] and multispectral images to produce 80 clusters at 98% convergence threshold. The mismatched pixels are assigned to their respective clusters after the analysis of their spectral value by classification feature set [3].

The K-means algorithm is a widely used unsupervised learning algorithm, which is a simple clustering technique in data mining. Clustering is a process of clustering similar data sets into groups as shown in Figure-1. Through an iterative converging process, this algorithm divides the data into K separate clusters and categorizes the data. The K-means algorithm categorizes data based on the K- value, where K stands for the number of clusters by calculating distance between each data point and the nearest center. The data points in the clusters and the K centroids are obtained by performing Euclidean distance measure [6].

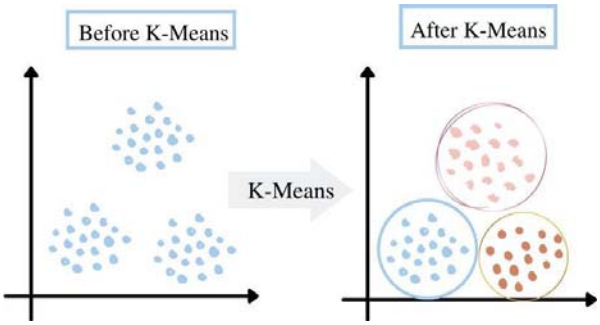


Figure-1. K-means clustering (Source: [24]).

Hyperion, Landsat 8 OLI, and Sentinel-2 data are classified using a k-nearest neighbor (KNN) algorithm. Support Vector Machine (SVM) is used for categorization of the AVIRIS-NG image. SVM is a machine learning technique that solves classification and regression problems using a supervised learning strategy. Mangrove species-level classification is essential for change detection, loss, and regeneration evaluation, recognition of climate change indicators, and implementation of government mitigation initiatives [8].

The researchers used supervised, unsupervised, parametric, non-parametric, object-oriented, per-pixel, sub-pixel, hard and soft classifiers to classify LULC changes. The K-mean text clustering technique, the multi-verse optimizer algorithm, feature selection and the enhanced krill herd algorithm are the important feature extraction methods that can be used for further processing of satellite images [10].

The images are sorted into numerous land use classifications using ERDAS Imagine 2013 software after image pre-processing. The process of classifying pixels into a finite number of separate classes or categories of data is performed based on their data file values [15].

Random Forest is an ensemble learning approach that uses multiple decision trees. Each decision tree is built using a fraction of the original training data and assessed using the remaining training attributes. New items are classified according to the class predicted by most of the trees [7, 20].

Water, vegetation, barren ground, and construction are the four major categories at the first classification level. Three pre-steps are involved in the classification of the first level. The first step is to construct the Normalized difference vegetation Index (NDVI), then the Normalized Difference Water Index (NDWI) and

finally image masking using these indices. Second-level classification and class combining are likewise the responsibility of ENVI. SVM and LULC maps are used for second-level categorization [4].

LULC classification methods and classifier techniques are extremely adaptable for extracting accurate LULC data from remote sensing images [13]. Classification is done on a small group of pixels, taking into account the spatial features of each pixel as they identify with one another [1].

1.3 Accuracy Assessment

Accuracy is the fraction of total samples correctly identified by the classifier [14, 16]. The accuracy is calculated using the accuracy score function as a percentage of correct predictions. In multi-label classification, the function returns the subset accuracy. If the entire set of premise labels for a sample closely matches the accurate set of labels, the subset accuracy is 1.0. The fraction of successful predictions across n samples [14] is defined as:

$$accuracy(x, y) = \frac{1}{n_{samples}} \sum_{i=0}^{n_{samples}-1} 1(y_i = x_i)$$

where,
1(xi) is the indicator function and
y_i is the predicted value of the i-th sample

The performance of a machine learning classifier can be measured using the confusion matrix. It is displayed in the form of a matrix. Actual and expected values are compared using the confusion matrix. The confusion matrix is a square matrix of order N, with N being the number of outputs or classifications [14].

In this research classification report for the supervised classifier includes precision, recall and Fbeta-score or measure.

1.3.1 Precision

It measures the classifier's ability to avoid classifying a negative sample as positive [14]. It is given by Equation (1).

$$tp / (tp + fp) \quad (1)$$

where,
tp is the number of true positives and
fp is the number of false positives, is the precision.

1.3.2 Recall

The capacity of a classifier to accurately detect all positive instances is measured by recall, which is a measure of its completeness [25]. It is calculated using the formula given in the Equation (2).

$$tp / (tp + fn) \quad (2)$$

1.3.3 Fbeta-score

The Fbeta-score is a weighted harmonic mean of precision and recall [21]. It is computed using the formula given in the Equation (3).

$$\text{Fbeta-score} = ((1+beta^2)*Precision*Recall)/(beta^2*Precision+Recall) \quad (3)$$

When the beta value is equal to 1, the precision and recall are perfectly balanced [23]. The number of actual instances of the class in the provided dataset is referred to as support [25].

A Kappa analysis (a multivariate discrete approach) and an error matrix are employed to evaluate the accuracy of the results. ERDAS IMAGINE software's accuracy assessment tool is employed [3].

All four mangrove classified images (Landsat 8 OLI, AVIRIS-NG, Sentinel-2 and Hyperion) are subjected to the accuracy assessment. On Lothian Island, the error matrix of each species group is employed to generate kappa statistics, user accuracy, producer accuracy and overall accuracy [8].

The reference data is used to verify the accuracy. Google Earth images are discovered to be the most commonly used reference data for evaluating classification accuracy. K (hat) statistics determine the accuracy of mapping distinct land use classes, as well as overall accuracy [10].

When it comes to remote sensing data, accuracy evaluation or validation is a crucial stage. The total accuracy of the classified image is determined by comparing how each pixel was arranged with the original land conditions obtained from the associated ground truth data. This software includes a tool for determining accuracy [15].

In addition to using the data effectively, the consumer of land-cover output has to know how reliable the outcome is. The accuracy assessment is a key stage in evaluating the classification process's result. The minimum degree of the accuracy of remote sensing data interpretation in identifying land use and LULC groups should be at least 85% [11].

Different factors influence the accuracy of the grouped guides, such as the size of the preparing test, type of the preparing test, classifier decision, size of the study area, etc., Understanding these factors helps in determining the appropriate category for a certain need [1].

2. STUDY AREA AND DATASET USED

2.1 Study Area

The location of the study area considered in this research is in the vicinity of GNITS, Shaikpet, Hyderabad, Telangana, India with latitude 78°23'57"-78°23'57" North and longitude 17°24’52"-17°24'34" East. The view of this study area from Google Earth Engine is shown in Figure-2.

Figure-2. Location of the study area (Credits: Google Earth Engine platform).

2.2 Dataset Used

Cloud free optical image collected by Sentinel-2B satellite on February 27, 2021 is used in this work. Sentinel is a Copernicus earth observation mission that collects optical images in a systematic way. Sentinel-2A and Sentinel-2B satellites used in this programme capture images of multi-spectral data with 13 bands. The resolution of these images ranges between 10m and 60m. Sentinel-2A and 2B satellites capture images with a revisit time of 10-days [20]. In this study, the near-infrared (NIR), red (R), blue (B), and green (G) bands are used for LULC classification. The wavelengths of the B, G, R, and NIR bands are 492.1, 559.0, 664.9, and 832.9 nanometers, respectively [17]. A Level-1C Sentinel-2B satellite image is collected from the United States Geological Survey (USGS) Earth explorer which provides free access to satellite images of different satellites. As this work is aimed to categorize LULC classes in the given input satellite image. The image selected for this study is one collected on 27th February 2021, so that in this month vegetation and water classes will be more available along with other classes such as barren and built-up classes than in the other months.

2.2.1 Data preparation

As the satellite images acquired are too large to be processed, sub-setting is performed to obtain an Area of Interest (AOI) covering the area in and around GNITS. Image sub-setting and band stacking is performed in the QGIS tool [22]. Sub-setting on each band is performed individually. Blue, green, red and near infrared bands are stacked after sub-setting the individual bands using the Semi-automatic classification plug-in in the QGIS tool as these bands are more frequently used for performing LULC as they give better results [3, 8, 19].

3. METHODOLOGY APPLIED

A Sentinel-2B satellite image acquired on February 27, 2021 is downloaded from the USGS Earth explorer. Pre-processing operations such as sub-setting, image stacking and image enhancement are performed. As an initial step, sub-setting is performed to obtain an AOI on the B, G, R and NIR bands of this image. Sub-setting is carried out on these bands, as the satellite image acquired by Sentinel-2B is too large to process and covers more geographical area than the AOI. Then the individual B, G, R and NIR bands of the AOI of this image are stacked. Image enhancement involves improving properties of the image. The image is enhanced by applying the PCA approach, this technique also reduces noise. [8]. To generate the training data red, green, blue and near infrared bands are used to apply the K-Means Clustering algorithm with K=4 which formed the four clusters based on the spectral similarity of classes. Figure-3 depicts the flow diagram.

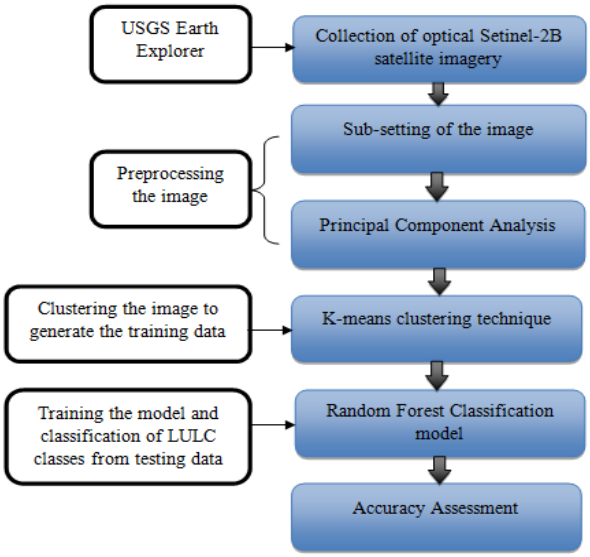


Figure-3. Flow diagram.

The output of K-means clustering is used as training data to build a Random Forest model to assign LULC classes to each pixel in a satellite image and the correctness of the Random Forest classifier's result is then verified using the testing dataset.

4. IMPLEMENTATION

4.1 Technologies Used

Sentinel-2B satellite image collected from USGS Earth Explorer is in GeoTiff format and it contains georeferencing information embedded in the GeoTiff file. Python contains a Rasterio library which supports reading data or images present in the Geotiff files. EarthPy is a Python library for manipulating geographical and remote sensing data. EarthPy has several Python packages including rasterio, geopandas and numpy.

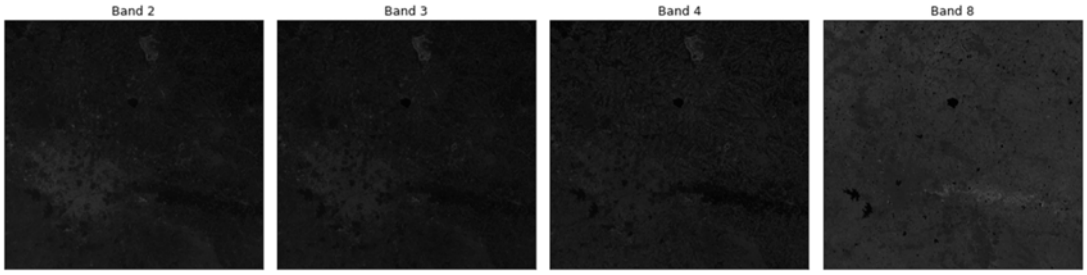
The Scikit-learn (Sklearn) package in python is one the efficient machine learning libraries. It gives a python framework for a variety of effective machine learning algorithms, such as categorization, modeling, grouping, and dimensionality reduction. This package, which is mostly written in Python and is built using Matplotlib, NumPy, and SciPy [14].

Quantum GIS (QGIS) is an open source Geographic Information System (GIS) that provides geographic data access, visualization, processing, and analysis. It can access vector data stored in a variety of formats, including file-based (ESRI Shape Files, KML, GML), geodatabases (e.g., PostgreSQL/PostGIs, ODBC, ESRI Personal GeoDatabase, SQLite), and network protocols (OPeNDAP, GeoJSON); raster data in one of over 40 formats supported by the underlying GDAL raster library (including NetCDF, HDF5, GeoTIFF, GRIB, and JPEG-2000). QGIS has a "plug-in" design that allows developers to create and utilize extensions to the application's basic capabilities [21].

5. RESULTS AND DISCUSSIONS

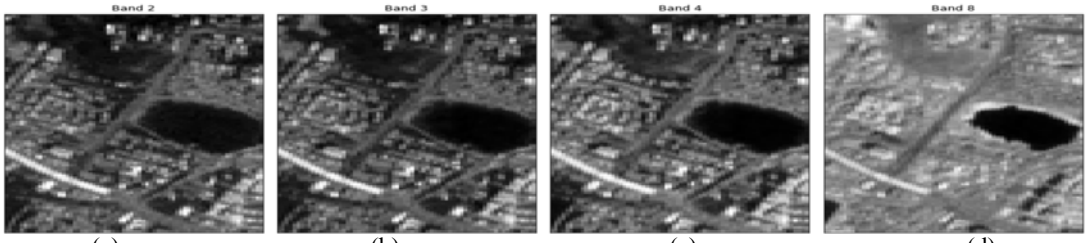
A Sentinel-2B satellite image acquired on February 27, 2021 is collected and preprocessing operations such as sub-setting and stacking the bands required for this research are performed using the QGIS tool. The spectral bands required for classifying are read using the rasterio library of python to perform the operations on the image. The PCA method is performed using sklearn. decomposition package in python [5]. The output components of PCA are given as inputs to K-Means Clustering algorithm which is implemented using sklearn. cluster package in python and using the sklearn. model selection package, the output of the K-means clustering technique is separated into training and testing data, and the training data is used to build a Random Forest model, which is implemented in Python using the sklearn. ensemble package.

5.1 Image Pre-Processing



(a) (b) (c) (d)

Figure-4 (a), (b), (c) & (d). B, G, R and NIR bands of Sentinel-2B image collected on February 27, 2021 respectively.



(a) (b) (c) (d)

Figure-5. B, G, R and NIR bands respectively after sub-setting.

The stacked input image is loaded into google drive and mounted on google colab. Google colab is an online version of Anaconda jupyter. Figure-4 shows the B, G, R and NIR bands of image collected by Sentinel-2B on February 27, 2021. Band 2 (refer Figure-4(a)) corresponds to B band, band 3 (refer Figure-4(b)) corresponds to G band, band 4 ((refer Figure 4(c)) corresponds to R band and band 8 ((refer Figure-4(d)) corresponds to NIR band in Sentinel 2B satellite image. The B, G, R and NIR bands of image collected by Sentinel-2B on February 27, 2021 after sub-setting to 102 rows x 68 columns are shown in Figure-5(a), (b), (C) & (d) respectively and the stacked image is shown in Figure-6.

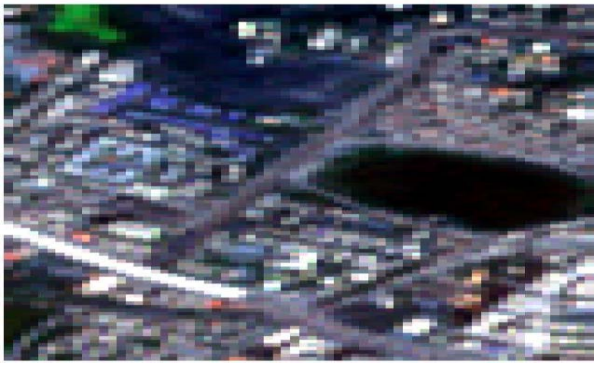
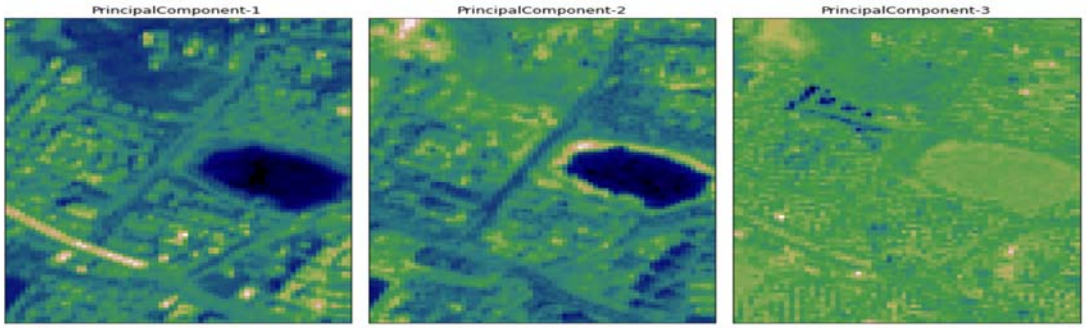


Figure-6. Stacked image of subsets of B, G, R and NIR bands of Sentinel-2B image collected on February 27, 2021.

PCA technique is applied to remove the noise from the input stacked Sentinel-2B image. The number of components of PCA is determined by visually inspecting the output of the K-means clustering algorithm with the Google Earth Engine view of the AOI (refer Figure-1).



(a) (b) (c)

Figure-7. Three principal components after applying PCA.

When the principal components chosen are three, the three output components are shown in Figure-7 (a), (b), & (c) respectively. The first component has variance of 0.8214 (refer Figure-7(a)) and the second component (refer Figure-7(b)) has variance of 0.1614 and third component (refer Figure-7(c)) has variance of 0.0130 and covered a total variance of 0.9958 [5].

5.2 Application of K-Means Clustering Technique

Figure-8 corresponds to the result obtained with the first principal component (refer Figure-7(a)) given as input to the K-means clustering model with K = 4 to obtain 4 clusters to identify 4 LULC classes. The first principal component (refer Figure-7(a)) with variance of 0.8214 is given as input to the K-means clustering method. When visually inspected with the Google Earth Engine view (Figure-1) it can be observed that clusters are not formed properly to identify LULC classes in the resultant image obtained from K-means clustering algorithm (refer Figure-8).

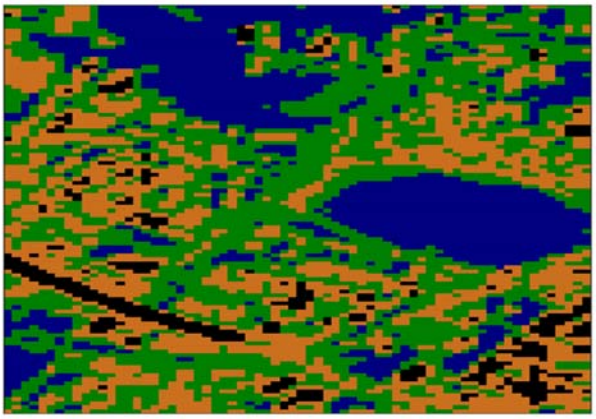


Figure-8. Results obtained from the K-means clustering algorithm with one principal component of PCA.

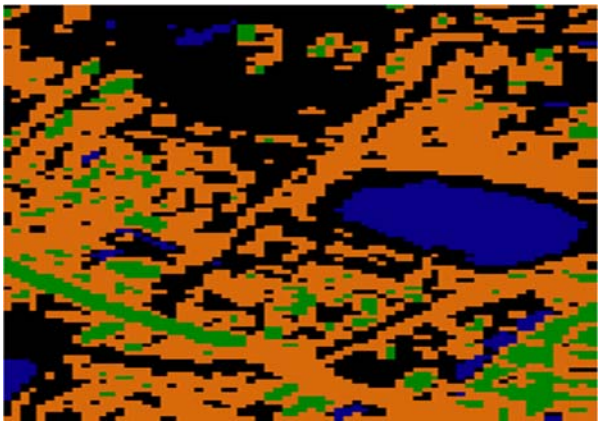


Figure-9. Results obtained from the K-means clustering algorithm with two principal components of PCA.

When the first two principal components (refer Figure-7(b)) are given as input to K-means clustering technique i.e. the first component with variance of 0.8214 and the second component with variance of 0.1614, with a total variance of 0.9828. The result of the K-means clustering algorithm when two principal components are chosen is shown in Figure-9. From the Figure-9, by performing visual inspection of the output of the K-means clustering method with the Google Earth Engine view (Figure-1), it can be observed that the clusters are formed in a better way to identify LULC classes (blue color representing water, vegetation in black color, built-up area in brown color and barren area in green color) than in Figure-9.

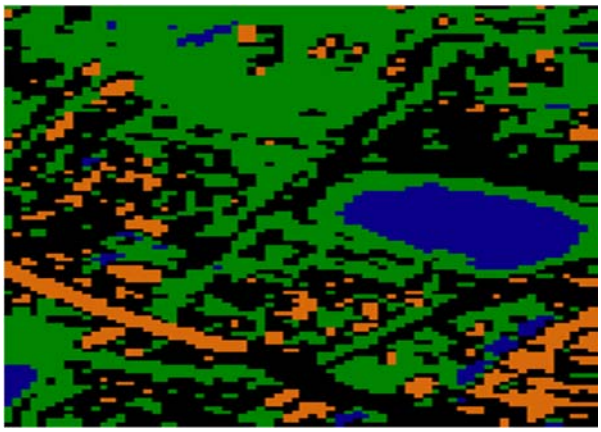


Figure-10. Results obtained from the K-means clustering algorithm with three principal components of PCA.

Figure-10 depicts the K-means clustering algorithm result when three principal components (refer Figure-7(c)) are chosen. By performing visual inspection of the output of the K-means clustering method (Figure-10) with the Google Earth Engine view (Figure-1) it can be observed that the clusters are formed in a better way to identify LULC classes in Figure-10 when compared to Figure-8 when one principal component is given as input to the K-means clustering algorithm. In Figure-10 the clustered pixels in blue color represents water, pixels in green color represents vegetation, pixels in black color represents built-up area and pixels in brown color represents barren area. When the number of principal components is reduced to two, vegetation and water areas are more precisely clustered (refer Figure-9). But when the principal components are three built-up and barren areas are clustered to some extent, while built-up and vegetation areas are not distinguished well (refer Figure-10) compared to when principal components are two (refer Figure-9).

5.3 Implementation of Random Forest

```
[[153,   0,   0,   0],
 [  0, 836,  24,   0],
 [  2,  34, 882,   2],
 [  0,  14,   0, 134]])
```

Figure-11. Confusion matrix of the Random Forest classifier model.

The output of the K-means clustering is used as LULC class for each pixel and all the pixels in the study area are split into training and testing data using sklearn package with ratio of 8:2 respectively. This splitting ratio is chosen because of the higher accuracy. Using the training data, a random forest classification model is trained initially. Then the test data is given as input to the Random forest classification model to classify into LULC classes. The class of the pixel in the test dataset is obtained with an accuracy of 78.34% when the maximum tree depth allowed is two. The accuracy of 82.96% is obtained when two is the maximum depth allowed and the accuracy of 87.60% is obtained when there is no limit on maximum tree depth. The confusion matrix and classification report are depicted in Figures 11 & 12 respectively.

In the confusion matrix (refer Figure-11), it can be observed that how many pixels of a particular class are classified by the model into each class. The classification report (refer Figure-12) of the Random Forest classifier model built contains precision which is the classifiers ability to avoid classifying a negative sample as positive, recall is the capacity of a classifier to accurately detect all positive instances and F-beta-score is referred as F- score, when beta=1 for the machine learning model built. Figure-13 shows the LULC classes of the research area obtained using the Random Forest classifier model.

	precision	recall	f1-score	support
0	0.99	1.00	0.99	153
1	0.99	1.00	0.99	860
2	1.00	0.99	0.99	920
3	0.99	0.99	0.99	148
accuracy			0.99	2081

Figure-12. Classification report of the Random Forest classifier model.

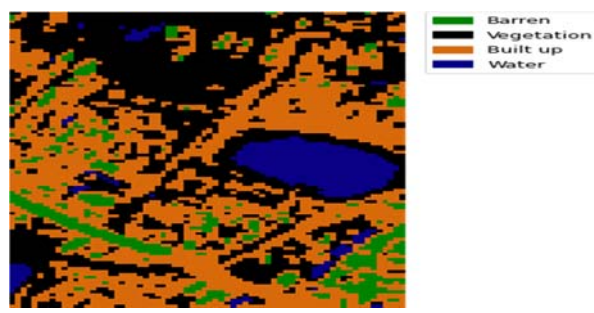


Figure-13. Classifications of study area using the Random Forest classifier model.

6. CONCLUSIONS

The work carried out in this paper discusses the classifying LULC using remotely sensed Sentinel-2B images. PCA is used to compress and eliminate noise after sub-setting and stacking B, G, R and NIR bands of Sentinel-2B image acquired on February 27, 2021 using QGIS Tool. To generate training data, the K-means clustering algorithm is used, which creates clusters based on spectral similarity. The Google Earth Engine perspective of the study area is used to perform a subjective evaluation of the clusters generated using K-means clustering and to determine the cluster's LULC class and also to decide the number of principal components to be used to produce clusters. The output of K Means clustering technique is used as class labels and the two principal components of bands of the image of the

study area are partitioned into training and testing sets to build the random forest classifier model. Random Forest classifier model with the different maximum depth that any tree in the model can grow have been built to get better results. The highest accuracy among the resulting models is 87.60%. The model built has been used to perform LULC classification of study areas. For future enhancements, it could be determined whether other bands of the study area's image can be used as input data to build more accurate LULC maps. The training data generated using K Means Clustering algorithm in this study can be used to build other machine learning models. The produced LULC classification in this paper can be used to detect changes in LULC in a specific area.

CONFLICT OF INTEREST

There is no conflict of interest.

ACKNOWLEDGEMENT

The authors would like to thank USGS Earth Explorer for providing Sentinel-2B data products.

REFERENCES

[1] Alshari., Eman & Gawali., Bharti. 2021. Development of Classification System for LULC Using Remote Sensing and GIS: Global Transitions Proceedings. 23: 8-17.

[2] Anusha N. and Bharathi B. 2020. Flood detection and flood mapping using multi-temporal synthetic aperture radar and optical data: The Egyptian Journal of Remote Sensing and Space Sciences. 23: 207-219.

[3] Azeem Uddin Siddiqui and Manish Kumar Jain. 2022. Change analysis in land use land cover due to surface mining in Jharia coalfield through Landsat time series data: Materials Today: Proceedings. 49: 3462-3468.

[4] Cavur, Mahmut, Duzgun., Sebnem, Kemec., Serkan, Demirkan and Doga. 2019. Land use and land cover classification of sentinel 2-A: St Petersburg case study. XLII-1/W2: 13-16.

[5] Dharani M. and Sreenivasulu G. 2019. Classification and Change Detection of Tirupati Urban Area using Erosion and Dilation Based PCA Transform: International Journal of Recent Technology and Engineering. 8.

[6] Hossain Md, Akhtar., Md. Nasim, Ahmad, R. Badlishah, Rahman and Mostafijur. 2019. A dynamic K-means clustering for data mining: Indonesian Journal of Electrical Engineering and Computer Science. 521. 10.11591: 521-526.

[7] Hutt, Christoph, Koppe., Wolfgang Miao., Yuxin Bareth and Georg. 2016. Best Accuracy Land Use/Land Cover (LULC) Classification to Derive Crop Types Using Multitemporal, Multisensor, and Multi-Polarization SAR Satellite Images: Remote Sensing. 8.

[8] Jyoti Prakash Hati., Sourav Samanta., Nilima Rani Chaube., Arundhati Misra., Sandip Giri., Niloy Pramanick., Kaushik Gupta., Sayani Datta Majumdar., Abhra Chanda., Anirban Mukhopadhyay and Sugata Hazra. 2021. Mangrove classification using airborne hyperspectral AVIRIS-NG and comparing with other spaceborne hyperspectral and multispectral data: The Egyptian Journal of Remote Sensing and Space Science. 24: 273-281.

[9] L. B. Prosper and Q. Guan. 2015. Analysis of land use and land cover change in Nadowli District, Ghana: 2015 23rd International Conference on Geoinformatics: 1-6.

[10] Loganathan., Agilandeeswari, Navin and Sam. 2020. Multispectral and hyperspectral images-based land use / land cover change prediction analysis: an extensive review, Multimedia Tools and Applications.

[11] Mohajane., Meriame, Essahlaoui., Ali, Oudija., Fatiha, el Hafyani., Mohammed, El Hmaidi., Abdellah, Ouali., Abdelhadi, Randazzo, Giovanni and Teodoro. 2018. Land Use/Land Cover (LULC) Using Landsat Data Series (MSS, TM, ETM+ and OLI) in Azrou Forest, in the Central Middle Atlas of Morocco: Environments. 5, 131. 10.3390/environments5120131.

[12] Ram Kumar Singh., Prafull Singh., Martin Drews., Pavan Kumar., Hukum Singh., Ajay Kumar Gupta., Himanshu Govil., Amarjeet Kaur and Manoj Kumar. 2021. A machine learning-based classification of LANDSAT images to map land use and land cover of India: Remote Sensing Applications: Society and Environment. 24.

[13] S. Hashimoto., T. Tadono., M. Onosato and M. Hori. 2013. Land use and land cover inference in large areas using multi - temporal optical satellite images: International Geoscience and Remote Sensing Symposium - IGARSS. 3333- 3336.

[14] Scikit-learn: Machine Learning in Python, Pedregosa *et al.* 2011. JMLR. 12: 2825-2830.

[15] Suneela T. and Mamatha G. 2016. Detection of Land Use and Land Cover Changes Using Remote Sensing and Geographical Information System (GIS) Techniques: International Journal of Electrical, Electronics and Data Communication. 4: 12.

[16] Vivekananda Gn., Swathi R. and Sujith. 2020. Multi-temporal image analysis for LULC classification and change detection: European Journal of Remote Sensing. 54: 1-11.

[17] Wikipedia contributors, Sentinel-2, Wikipedia, The Free Encyclopedia, Available at: https://en.wikipedia.org/w/index.php?title=Sentinel-2&oldid=1066678044, (2022).

[18] Xu H., Yao S., Li Q and Ye Z. 2020. An Improved K-means Clustering Algorithm. IEEE 5th International Symposium on Smart and Wireless Systems within the Conferences on Intelligent Data Acquisition and Advanced Computing Systems.

[19] Yu, Zhiqi., Di, Liping., Yang, Ruixing., Tang, Junmei., Lin Li., Zhang Chen., Rahman M., Zhao Haoteng., Gaigalas Juozas., Yu Eugene and Sun Ziheng. 2019. Selection of Landsat 8 OLI Band Combinations for Land Use and Land Cover Classification, Agro-Geoinformatics. 1-5.

[20] Zhang Fang and Yang Xiaojun. 2020. Improving land cover classification in an urbanized coastal area by random forests: The role of variable selection.Remote sensing of Environment. 251.

[21] https://medium.com/@douglaspsteen/beyond-the-f-1-score-a-look-at-the-f-beta-score-3743ac2ef6e3.

[22] https://old.dataone.org/software-tools/quantum-gis-qgis.

[23] https://thecleverprogrammer.com/2021/07/10/f-beta-score-in-machine-learning.

[24] https://www.javatpoint.com/k-means-clustering-algorithm-in-machine-learning.

[25] https://www.scikit-yb.org/en/latest/api/classifier/classification_report.html.