Classification of Multi-spectral Landsat Satellite Imagery Dataset using Feature Extraction Methods

Saurabh Kumar, Shwetank

Abstract: The feature extraction of multi-spectral Landsat satellite imagery Dataset is essential for vegetation monitoring, urban planning, change assessment, and other land-use applications. The spatial information provided by Remote sensing satellite imagery data is helpful for planning and decision-making policies. In the present study, classify the features of the multi-spectral Landsat satellite imagery dataset in different periods using the feature extraction method, and is produced the spatial maps of the study area. The study is to analyze the appropriate method of feature extraction for classifying the orchards, vegetation, rangeland, agricultural land, wetland, water body, and urban land using multi-temporal satellite dataset. In this study, use the three feature extraction methods are support vector machine (SVM), minimum distance (MD), and Maximum likelihood classifier (MLC) for supervised pixel-based classification using medium resolution (30 m) satellite dataset. The accuracy of feature extraction method is performed by the MLC (86.29% and 93% in the year 2003 and 2017) and SVM (86.37% and 90% in the year 2003 and 2017). The result of the presented study shows MLC and SVM classifier performs similar results but better than MD classifier for land-use/cover features classification. The classified spatial maps provide the essential spatial information for land-use changes occurred during the last 15 years (2003 to 2017).

Keywords: feature extraction, remote sensing, spectral, spatial maps.

I. INTRODUCTION

The remote sensing (RS) feature extraction method is to explore the potential of multispectral satellite imagery data and produce the spatial maps. The spatial maps are essential for monitoring and planning of land-use [1] of the study area. The satellite imagery data play a vital role in feature extraction from the earth's surface and provides spatial and temporal information of land cover [2]. The RS shows an essential role in urban sprawl and growth and its impact on vegetation and forest [3].

Revised Manuscript Received on February 28, 2020.
* Correspondence Author
 Saurabh Kumar*, Ph.D. Scholar, Department Of Computer Science, Gurukula Kangri Vishwavidyalaya, Haridwar, Uttarakhand, India. Email: rs.saurabhkumar@gkv.ac.in, saurabhpathakkanpur@gmail.com
 Shwetank, Assistant Professor, Department Of Computer Science, Gurukula Kangri Vishwavidyalaya, Haridwar, Uttarakhand, India. Email: Shwetank.arya@gmail.com

Numerous researchers had applied the multispectral Landsat satellite data to monitoring and observations in local and global for land use applications because USGS freely provides it in digital format and contains a wide range of spectral bands [4-5]. The multispectral Landsat dataset has the potential to mapping, observations, monitoring, and change assessment in various fields of vegetation [6], agriculture [7], forestry [8], and urbanization [9]. The RS techniques are beneficial in environment monitoring [10] and change predictions. The RS and geographical information system (GIS) tools are essential for spatial analysis and change assessment in land use applications (vegetation mapping, crops cultivation monitoring, urban mapping, etc.). The study is to analyze the appropriate supervised classification method for feature extraction from remote sensing satellite images.

II. STUDY AREA

The Haridwar region selected as a study area is situated in Uttarakhand state of the south-western position between 29.58º latitude and 78.13º longitude. Haridwar is the gate of holy river Ganga and a famous Hindu pilgrimage. The weather of the study region is in summers around 25ºC to 44ºC and winters -2ºC to 24ºC, yearly regular rainfall 1174.3 mm. Haridwar district has contained three subdivisions (viz. Haridwar, Laksar, and Roorkee), and covered 2,360 km² area. It has six administrative blocks and 610 villages, with 18, 90, 422 population according to the 2011 census. Haridwar region is a vital place for horticultural fruits and crop cultivation in Uttarakhand. The QGIS 2.18 (open-source tool) is used to prepare the spatial maps of the study region and show in figure 1.
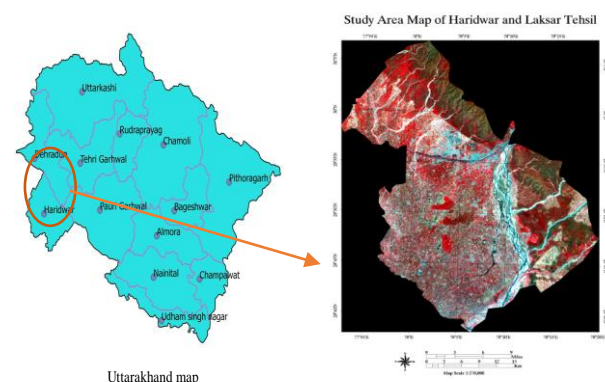

Fig. 1. Map of study area

III. DATASET

In the proposed study, Landsat-8 OLI and Landsat-7 ETM+ satellite imagery dataset used as primary data, and secondary data are used as reference data for accuracy assessment of classified dataset. The secondary dataset has collected information (in the form of ground control points (GCP)) by ground survey using a hand-held Garmin 60 GPS device. The topographic map of the Haridwar region is essential for accurate assessment of primary data acquired from the survey of India Dehradun at the 1/25,000 scale. The secondary information has led to the approval of the correctness of information despite the whole study depend on the primary dataset. All essential information about the primary dataset has acquired from the United States Geological Survey (USGS) mentioned in table 1.

Table 1. Detailed of acquired Landsat 8 OLI satellite imagery data used for feature extraction.

Name of dataset	Source of data	Acquirement date	Spatial resolution	Band	UTM Zone/ Datum	Wavelength (μm)
Landsat-7 ETM+	USGS	07/03/2003	30 and 15 m (B8)	B1-Blue, B2-Green, B3-Red, B4-NIR, B5-SWIR, B6-TIR, B7-SWIR and B8-PAN	44/WGS 84	B1(0.45-0.52), B2(0.52-0.60), B3(0.63-0.69), B4(0.77-0.90), B5(1.55-1.75), B6(10.40-12.50), B7(2.09-2.35) and B8(0.52-0.90)
Landsat-8 OLI	USGS	08/05/2017	30 m (B1-B7),15 m (B8)and 100 m (B10-B11)	B1-Coastal Aerosol B2-Blue, B3-Green, B4-Red, B5-NIR, B6-SWIR, B7-SWIR, B8-PAN, B9-Cirrus, B10-TIRS 1 and B11-TIRS 2	44/WG S84	B1(0.43-0.45), B2(0.45-0.51), B3(0.53-0.59), B4(0.63-0.67), B5(0.85-0.87), B6(1.55-1.65), B7(2.10-2.29), B8(0.50-0.67), B9(1.36-1.38), B10(10.60-11.19) and B11(11.50-12.51)

IV. METHODOLOGY

The proposed methodology for feature extraction is used multi-spectral Landsat-8 OLI dataset as a primary data and ground survey, topographic map, and google earth images as secondary data. The preprocessing of a primary dataset is an essential part of accurate feature extraction and analysis.
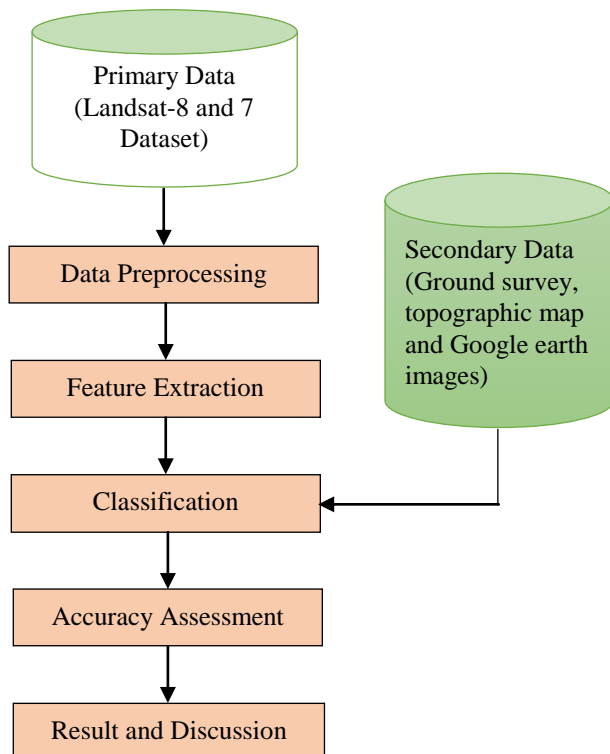

Fig. 2. Methodology for feature extraction.

The satellite images are preprocessed using the Geometric correction and Atmospheric correction method.

The Geometric correction method is used for eliminating the geometric distortion. Since two different periods (2003 and 2017) images are used for feature extraction, so that image registration is essential for the multi-temporal study. The Landsat-8 image (2017) is used as a reference image for registered to another image.

The atmospheric correction method is used to reduce the atmospheric disturbance and recover the surface reflectance from images by removing atmospheric effects. The atmospheric disturbance problems are water vapour and scattering (due to molecules and aerosols), absorption (due to CO_2, O_2, O_3) effects in the atmosphere. Atmospheric correction is most important when extracting features based on the spectral response of images picked up under different circumstances. The overall process of feature extraction is show in the figure 2.

A. Feature extraction from multi-spectral Landsat dataset

The land of the study area is covered various types of land-use, such as orchards, vegetation, open mixed forest, dense mixed forest, agricultural land, rangeland, watershed, water body, and urban. The land cover is distributed into seven classes according to the spectral response of reflected energy with the help of Google earth images and topographic maps. All details of land-use classes are described in table 2.

Table 2. Details of land cover classes for feature extraction.

Sr. No.	Type of classes	Color of classes	Details of classes
1.	Orchards	Green	Various types of horticulture fruit crops mainly mango, litchi, etc.
2.	Vegetation	Sea Green	Dense mixed forest, Poplar plants, Eucalyptus plants and Hills land.
3.	Rangeland	Cyan	Woodland, shrub land, open mixed forest, grassland and wetland
4.	Agricultural land	Yellow	Farmland, crops land (wheat, rice, sugarcane, etc.) and pasture
5.	Urban land	Red	Urban-Inhabited, commercial area, rural community, build-up Inhabited region, Industrial zones and roads
6.	Water body	Blue	River, waterway and pounds
7.	Watershed	White	Watershed land

The total area of land-use classes in the study area is covered 1, 20,136.05 ha. The maximum land cover is covered by agricultural land and minimum land covered by the water body. The orchards are also playing an essential role in food security and the economy of the Haridwar region. The vegetation and rangeland of the Haridwar region contain dense mixed forest, hills plant, open mixed forest, shrubs wetland plants, grasses, etc.

B. Accuracy assessment

Accuracy assessment measures the accuracy of all applied feature extraction methods individually. In this study, the Post classification method has measured the accuracy of the classifier using the confusion matrix show in table 3 to 8. The confusion matrix contains spatial information of known and predicted class variables. The Confusion matrix elements (i and j) are no. Samples with predicted class i and known class j. the diagonal elements represent accurately classified observations. Accuracy of classification is depended on various factors such as the resolution of satellite data, preprocessing, data sampling, secondary data, and appropriate selection of feature extraction methods. The ground survey, topographic map, and google earth images play a significant role in accuracy assessment.
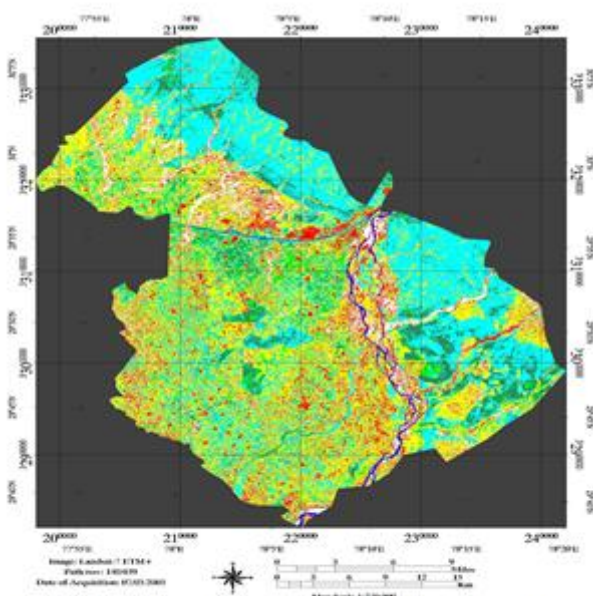
V. RESULT

multispectral Landsat satellite imagery data as compare to the minimum distance method. The classifier classifies the geographic feature such as orchards, vegetation, rangeland, agricultural land, watershed, and water body from multitemporal images (in the year 2003 and 2017) show in classified maps (figure 3 to 8).

Fig.: 4 Landsat-7 (2003) Classified image using MLC

Fig.: 3 Landsat-7 (2003) Classified image using MD

Fig.: 5 Landsat-8 (2003) Classified image using SVM

In this study, feature extraction classifier MLC and SVM both are classified with reasonable accuracy to multispectral and

Retrieval Number: D1589029420/2020©BEIESP
DOI:10.35940/ijrte.D1589.018520
Journal Website: www.ijitee.org

1868

Published By:
Blue Eyes Intelligence Engineering
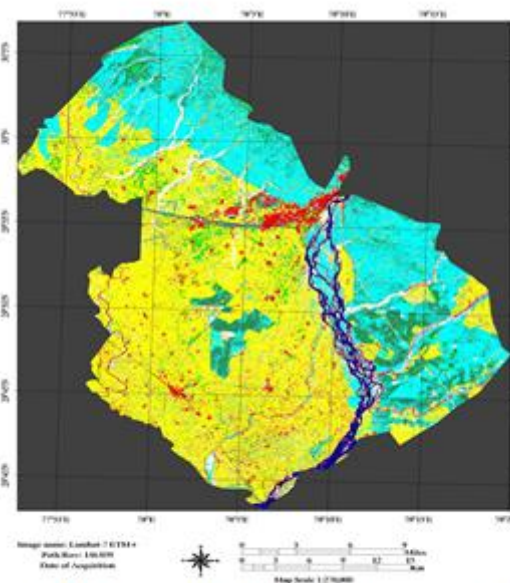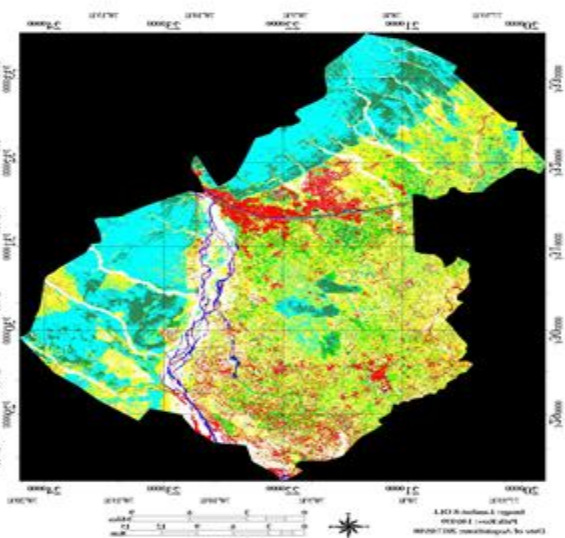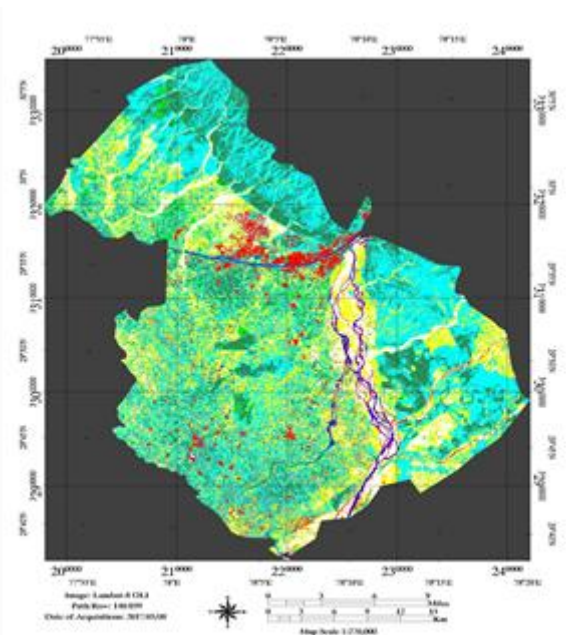*& Sciences Publication*

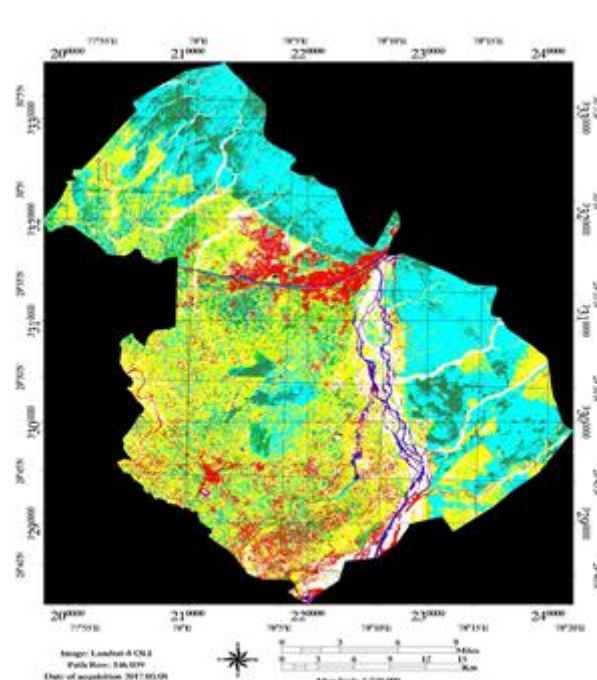Fig.: 6 Landsat-8 (2017) Classified image using MD
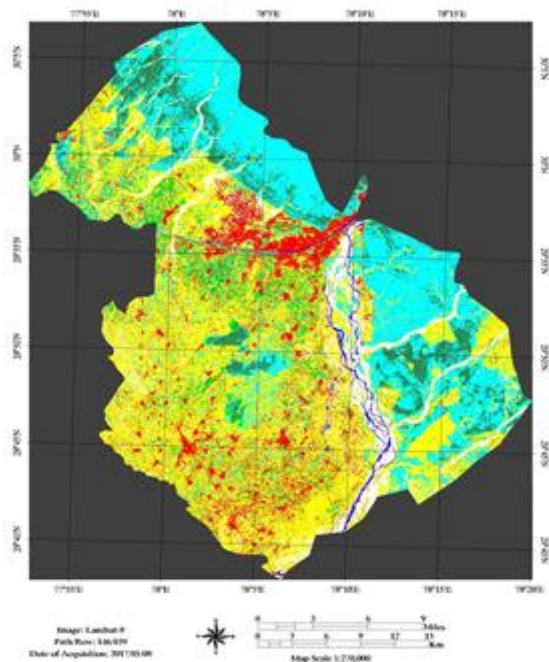

Fig.: 8 Landsat-8 (2017) Classified image using SVM

Fig.: 7 Landsat-8 (2017) Classified image using MLC

Each class in the classified spatial map is shown with a unique color. The classified spatial maps are show changes in land use during the last 15 years (2003 to 2017).

MLC and SVM are two standard classifier methods in machine learning. MLC classifier estimates the parameters of the probabilistic model based on Bayesian theory, while SVM, in this case, is based on an optimized nonparametric method. Recently, it was found that in the probabilistic modeling learning process, SVM is equivalent to MLC in some cases.

The accuracy of MLC classifier is 86.29% and 93% with kappa 0.83, 0.91 in 2003 and 2017 respectively, and for SVM classifier the accuracy is 86.37% with kappa 0.83, 90% with kappa 0.87 in 2003 and 2017 respectively and for MD classifier the accuracy is 71.63% with kappa 0.66, 78.56% with kappa 0.73 in 2003 2017. The accuracy of SVM and MLC classifier in this study shows perform better accurate result than minimum distance classifier for land-use classification and analysis.

Table 3. Confusion matrix of land-use classes classify by the MD in 2003

Class	Ground Truth Data (%)							
	Orchards	Vegetation	Rangeland	Agri. Land	Urban land	Water body	Watershed	Total
Orchards	**44.98**	20.42	9.21	5.16	0.00	0.00	0.85	**9.40**
Vegetation	44.60	**78.82**	1.67	0.10	0.00	0.00	0.00	**15.98**
Rangeland	7.04	0.76	**77.74**	19.37	1.38	0.38	3.19	**17.98**
Agri. Land	1.39	0.00	10.02	**48.27**	14.35	0.00	6.93	**13.47**
Urban land	0.00	0.00	0.59	11.73	**79.99**	6.32	15.89	**20.73**
Water body	0.00	0.00	0.72	0.19	0.00	**91.76**	0.00	**7.62**
Watershed	0.00	0.00	0.05	15.18	4.23	1.53	**73.14**	**14.83**
Total	100	100	100	100	100	100	100	**100**
Overall accuracy	71.63%							
Kc Value	0.66							

Table 4. Confusion matrix of Landuse classes classify by the MLC in 2003.

Class	Ground Truth Data (%)							
	Orchards	Vegetation	Range Land	Agri. land	Urban land	Water body	Watershed	Total
Orchards	85.73	7.46	0.63	3.75	0.12	0.00	0.10	8.64
Vegetation	11.50	90.37	2.12	0.97	0.00	0.00	0.35	15.43
Rangeland	0.10	1.46	87.86	10.71	0.04	0.00	1.15	17.29
Agri. Land	2.18	0.71	8.17	76.69	5.77	0.00	9.72	16.66
Urban land	0.00	0.00	0.72	3.11	90.55	2.39	8.52	19.98
Water body	0.00	0.00	0.00	0.24	0.00	96.17	0.00	7.86
Watershed	0.00	0.00	0.50	4.53	3.52	1.44	86.17	14.15
Total	100	100	100	100	100	100	100	100
Overall accuracy	86.29%							
Kc Value	0.83							

Table 5. Confusion matrix of Landuse classes classify by the SVM in 2003.

Class	Ground Truth Data (%)							
	Orchards	Vegetation	Range Land	Agri. land	Urban land	Water body	Watershed	Total
Orchards	75.02	7.01	0.23	2.63	0.08	0.00	0.20	7.48
Vegetation	19.82	90.22	2.93	0.97	0.00	0.00	0.20	16.18
Rangeland	0.79	2.37	89.93	10.90	0.08	0.00	1.59	17.95
Agri. Land	4.16	0.40	6.05	77.96	5.69	0.00	7.22	16.16
Urban land	0.10	0.00	0.77	4.14	91.18	1.63	9.87	20.43
Water body	0.00	0.00	0.00	0.19	0.24	97.70	0.00	8.02
Watershed	0.10	0.00	0.09	3.21	2.73	0.67	80.92	13.78
Total	100	100	100	100	100	100	100	100
Overall accuracy	86.37 %							
Kc Value	0.83							

Table 6. Confusion matrix of Landuse classes classify by the MD in 2017.

Class	Ground Truth Data (%)							
	Orchards	Vegetation	Range land	Agri. Land	Urban land	Water body	Watershed	Total
Orchards	85.44	4.91	1.23	0.14	0.00	0.00	0.00	6.70
Vegetation	14.56	89.88	15.43	1.66	0.17	0.00	0.05	26.22
Rangeland	0.00	5.20	66.70	22.96	16.02	0.00	0.16	16.33
Agri. Land	0.00	0.00	16.56	56.71	4.89	0.00	9.24	11.67
Urban land	0.00	0.00	0.00	0.69	78.41	0.72	12.70	10.32
Water body	0.00	0.00	0.09	0.00	0.00	99.28	0.05	6.07
Watershed	0.00	0.00	0.00	17.84	0.51	0.00	77.79	22.68
Total	100	100	100	100	100	100	100	100
Overall accuracy	78.56%							
Kc Value	0.73							

Table 7. Confusion matrix of Landuse classes classify by the MLC in 2017.

Class	Ground Truth Data (%)							
	Orchards	Vegetation	Range land	Agri. Land	Urban land	Water body	Watershed	Total
Orchards	98.57	2.57	0.00	0.14	0.00	0.00	0.00	6.72
Vegetation	0.72	93.39	1.23	0.00	0.00	0.00	0.00	23.67
Rangeland	0.24	3.98	93.43	6.36	0.17	0.00	0.00	17.33
Agri. Land	0.24	0.06	4.91	88.24	4.89	0.00	6.17	12.30
Urban land	0.24	0.00	0.00	2.07	93.59	0.00	2.25	8.97
Water body	0.00	0.00	0.00	0.00	0.00	98.80	0.00	6.01
Watershed	0.00	0.00	0.44	3.18	1.35	1.20	91.48	25.00
Total	100	100	100	100	100	100	100	100
Overall accuracy	93.00							
Kc Value	0.91							

Table 8. Confusion matrix of Landuse classes classify by the SVM in 2017.

Class	Ground Truth Data (%)							
	Orchards	Vegetation	Rangeland	Agri. Land	Urban land	Water body	Watershed	Total
Orchards	89.02	0.94	0.88	0.28	0.00	0.00	0.00	5.88
Vegetation	10.50	95.15	6.22	0.28	0.00	0.00	0.00	25.57
Rangeland	0.00	3.92	86.33	13.14	0.17	0.00	0.00	16.83
Agri. land	0.48	0.00	6.57	76.07	6.75	0.00	4.67	11.03
Urban land	0.00	0.00	0.00	0.14	86.68	0.48	3.13	8.42
Water body	0.00	0.00	0.00	0.00	0.00	99.52	0.05	6.07
Watershed	0.00	0.00	0.00	10.10	6.41	0.00	92.14	26.20
Total	100	100	100	100	100	100	100	100
Overall accuracy	90.00%							
Kc Value	0.87							

VI. CONCLUSION

In the present study, the RS feature extraction method is explored the geographic feature from the multispectral Landsat 8 OLI and Landsat-7 ETM satellite dataset in the form of spatial maps. The spatial maps are provided the essential spatial information for vegetation mapping, crop monitoring, and urban mapping and planning. The MLC and SVM Classifier classifies the land-use features from the multi-temporal images (in the year 2003 and 2017). The result of this study shows the accuracy of MLC (86.29% and 93% in the year 2003 and 2017) and SVM (86.37% and 90% in the year 2003 and 2017) classifier perform better accurate result than minimum distance classifier (71.63% and 78.56 % in the year 2003 and 2017) for the land-use classification and change assessment.

ACKNOWLEDGMENT

REFERENCES

1. Gao Yan, J. -F. Mas , B. H. P. Maathuis , Zhang Xiangmin & P. M. Van Dijk, "Comparison of pixel-based and object-oriented image classification approaches—a case study in a coal fire area, Wuda, Inner Mongolia, China", International Journal of Remote Sensing, vol. 27, no. 18, pp. 4039–4055, 2007.
2. Peng Gong, et al., "Finer resolution observation and monitoring of global land cover: first mapping results with Landsat TM and ETM+ data", International Journal of Remote Sensing, vol. 34, no.7, pp. 2607–2654, 2013.
3. J.S. Rawat, Manish Kumar, "Monitoring land use/cover change using RS and GIS techniques: A case study of Hawalbagh block, district Almora, Uttarakhand, India" The Egyptian Journal of RS and Space Sciences, no. 18, pp. 77–84, 2015.
4. Jingxiao Zhang, Li Jia, Massimo Menenti and Guangcheng, "Glacier Facies Mapping Using a Machine-Learning Algorithm: The Parlung Zangbo Basin Case Study", Remote Sensing, vol. 11:452, 2019.
5. Can Bulent Karakus, Orhan Cerit and Kaan Sevki Kavak, "Determination of land use/ cover changes and land use potentials of Sivas city and its surroundings using GIS and RS", Procedia Earth and Planetary Science, vol. 15, pp. 454-461, 2015.
6. DU Peijun, LI Xingli, CAO Wen, LUO Yan, ZHANG Huapeng, "Monitoring urban land cover and vegetation change by multi-temporal RS information", Mining Science and Technology, vol. 20, pp. 0922–0932, 2010.
7. Waldner F., De Abelleyra D., Verón S.R., Zhang M., Wu B., Plotnikov D., Bartalev S., Lavreniuk M., Skakun S., Kussul N., et al., "Towards a set of agrosystem-specific cropland mapping methods to address the global cropland diversity", International Journal of Remote Sensing, vol. 37, no. 14, pp. 3196–3231, 2016.
8. C. Sudhakar Reddy, C. S. Jha, P. G. Diwakar & V. K. Dadhwal, "Nationwide classification of forest types of India using RS and GIS", Environ Monit Assess Environ, vol. 187, no. 12: 777, 2015.
9. Rajesh Bahadur Thapa and Yuji Murayama, "Examining Spatiotemporal Urbanization Patterns in Kathmandu Valley, Nepal: Remote Sensing and Spatial Metrics Approaches", Remote Sensing, vol. 1, pp. 534-556, 2009.
10. Azzari G., Jain M., Lobell D.B., "Towards fine resolution global maps of crop yields: Testing multiple methods and satellites in three countries", Remote Sensing Environment, vol. 202, pp. 129–141, 2017.

AUTHORS PROFILE

Saurabh Kumar, Ph.D. scholar, department of computer science, Gurukula Kangri Vishwavidyalaya, Haridwar. The area of interest in Remote sensing, GIS and Digital image processing.

Dr. Shwetank, assistant professor, department of computer science, Gurukula Kangri Vishwavidyalaya, Haridwar. The area of interest in Remote sensing, GIS, Digital image processing and Data mining. He has 15 year experience in teaching and research, and published many research paper in the field of Digital image processing and Remote sensing.

Retrieval Number: D1589029420/2020©BEIESP
DOI:10.35940/ijrte.D1589.018520
Journal Website: www.ijitee.org

1871

Published By:
Blue Eyes Intelligence Engineering
& Sciences Publication