

A COMPARATIVE STUDY OF ADVANCED LAND USE/LAND COVER CLASSIFICATION ALGORITHMS USING SENTINEL – 2 DATA

K. Nivedita Priyadarshini ^{1,*}, Minakshi Kumar ², S. Abdul Rahaman ¹, S. NitheshNirmal ¹

¹ Dept. of Geography, Bharathidasan University, Tamil Nadu - nivi.darshini@yahoo.com, (abdulatgeo, nirmalgsarath) @gmail.com

² Indian Institute of Remote Sensing (ISRO), Dehradun, Uttarakhand – minakshi@iirs.gov.in

Commission V, SS: Emerging Trends in Remote Sensing

KEY WORDS: Land Use/Land Cover, Classification Algorithms, Sentinel-2A, Supervised, Unsupervised, Accuracy Assessment

ABSTRACT:

Land Use/ Land Cover (LU/LC) is a major driving phenomenon of distributed ecosystems and its functioning. Interpretation of remote sensor data acquired from satellites requires enhancement through classification in order to attain better results. Classification of satellite products provides detailed information about the existing landscape that can also be analyzed on temporal basis. Image processing techniques acts as a platform for analysis of raw data using supervised and unsupervised classification algorithms. Classification comprises two broad ranges in which, the analyst specifies the classes by defining the training sites called supervised classification where as automatically clustering of pixels to the defined number of classes namely the unsupervised classification. This study attempts to perform the LU/LC classification for Paonta Sahib region of Himachal Pradesh which is a major industrial belt. The data obtained from Sentinel 2A, from which the stacked bands of 10m resolution are only used. Various classification algorithms such as Minimum Distance, Maximum Likelihood, Parallelepiped and Support Vector Machine (SVM) of supervised classifiers and ISO Data, K-Means of unsupervised classifiers are applied. Using the applied classification results, accuracy assessment is estimated and compared. Of these applied methods, the classification method, maximum likelihood provides highest accuracy and is considered to be the best for LU/LC classification using Sentinel-2A data.

1. INTRODUCTION

1.1 Land Use / Land Cover in Remote Sensing

Mapping of LU/LC features has become the most applicable factor in various fields of geospatial technology (Melesse et al., 2007). Analysis of LU/LC in change detection purposes shows the technological capacity and improvement over years (Madhu et al., 2017). Using the existing land cover data, further studies on construction of urban patterns, industries, transport lines, conservation of water and other natural resources shall be monitored and maintained (Amna et al., 2015; Rahaman et al., 2017; Nitheshnirmal et al., 2017).

1.2 Classification of Land Use/ Land Cover information

Remote Sensing data are raw imagery that could be interpreted only through classification in order to obtain relevant information. The information extracted provides the existing land cover and land use patterns (Topaloğlu et al., 2016). Classification of raw satellite imagery through image processing techniques provides accurate details about the landscape. Especially, modelling environmental issues using classification algorithms is used at a wider range (Al-Ahmadi et al., 2008). There has been numerous studies conducted on the classification accuracies of popularly available Landsat images (TM, ETM+, OLI) with a spatial resolution of 30 m (Congalton, 1991; Huang et al., 2010; Rwanga and Ndambuki, 2017).

In recent times, Sentinel-2A provides data with higher spatial resolution than Landsat images which made researchers around the globe to opt Sentinel-2A data for their land use/land cover classification in various studies such as wetland monitoring (Kaplan and Avdan, 2017), crop and tree species classification (Immitzer et al., 2016), urban sprawl (Lefebvre et al., 2016), urban green space analysis (Kopecká et al., 2017) and many other studies where Sentinel-2A data is used to generate various thematic layers. Therefore, it is essential to check which among many classification algorithm best suits for the land use/land cover classification of Sentinel-2A data. There has been studies comparing the classification accuracies of Landsat-8 and Sentinel-2A with SVM and MLC classification algorithms (Topaloğlu et al., 2016; Sekertekin et al., 2017). The abovesaid studies concluded that Sentinel-2A have more accuracy than Landsat-8 by comparing both datasets with SVM and MLC algorithms, but there has been no detailed study on other supervised and unsupervised classification algorithms such as parallelepiped, minimum distance, ISO Data and K-Means for Sentinel-2A data. Therefore, this study is conducted to fill this research gap which aims to classify the Sentinel 2A data using the ENVI classification algorithms and to compare their respective classification accuracies to conclude which classification algorithm is best suited for Sentinel-2A data.

2. STUDY AREA AND DATASETS

2.1 Study area description

This study aims to perform the LU/LC classification for Paonta Sahib region of Himachal Pradesh. It lies between 30.4453° N latitude and 77.6021° E longitude. Paonta Sahib is a sub division under the Sirmour district. The study area is one among the lowest administrative unit but considered useful to the

* Corresponding author

planners who formulate micro level developmental plans. Hence it was termed as the community development block by the Census of India in the year 1991. The total area covered is 762.91 Sq. Km and lies at the foothills of the Siwaliks. The dense mountainous ranges are found to the north of river Yamuna which runs across Paonta Sahib. Prospective industries are present along the banks of the river. The north western

terrain is dotted with dense vegetation and reserved forests. The Paonta Sahib region is heavily concentrated with linear and nucleated settlements along the roads and river. Patches of cultivated lands are contributed to agriculture. Since the study area exhibits a diverse landuse/landcover pattern, an attempt has been made to identify it using various classification methods.

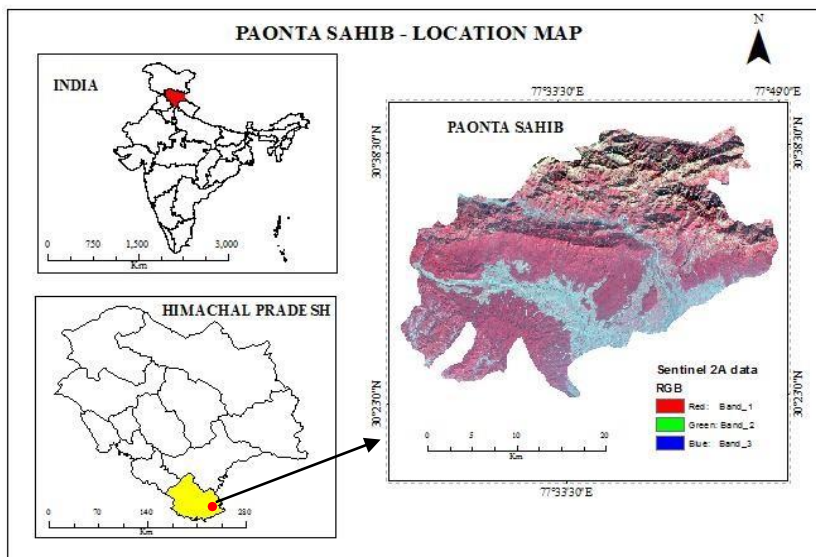


Figure 1. Paonta Sahib – Location map

2.2 Sentinel 2 data description

Sentinel 2 was developed by ESA and consisted of two identical satellites Sentinel 2A and 2B. The multi spectral data of Sentinel 2A has a total number of 13 bands and was launched on June 23rd 2015. The spatial resolution of Sentinel 2A varies as 10m, 20m and 60m for different bands. The Sentinel 2 carries a single Multi Spectral Instrument with 13 spectral channels whose band specifications are given in Table 2.

Sentinel-2 Bands	Central Wavelength (µm)	Resolution (m)
Band 1 - Coastal aerosol	0.443	60
Band 2 - Blue	0.490	10
Band 3 - Green	0.560	10
Band 4 - Red	0.665	10
Band 5 - Vegetation Red Edge	0.705	20
Band 6 - Vegetation Red Edge	0.740	20
Band 7 - Vegetation Red Edge	0.783	20
Band 8 - NIR	0.842	10
Band 8A - Vegetation Red Edge	0.865	20
Band 9 - Water vapour	0.945	60
Band 10 - SWIR - Cirrus	1.375	60
Band 11 - SWIR	1.610	20
Band 12 - SWIR	2.190	20

Table 2. Sentinel 2A band specifications

3. METHODOLOGY

3.1 Method opted

The overall workflow in Figure 3, shows the processing of raw remote sensor data into classified data using various classification algorithms. The steps involved denote pre-

processing satellite data like resampling, layer stacking, masking, applying classification, accuracy assessment.

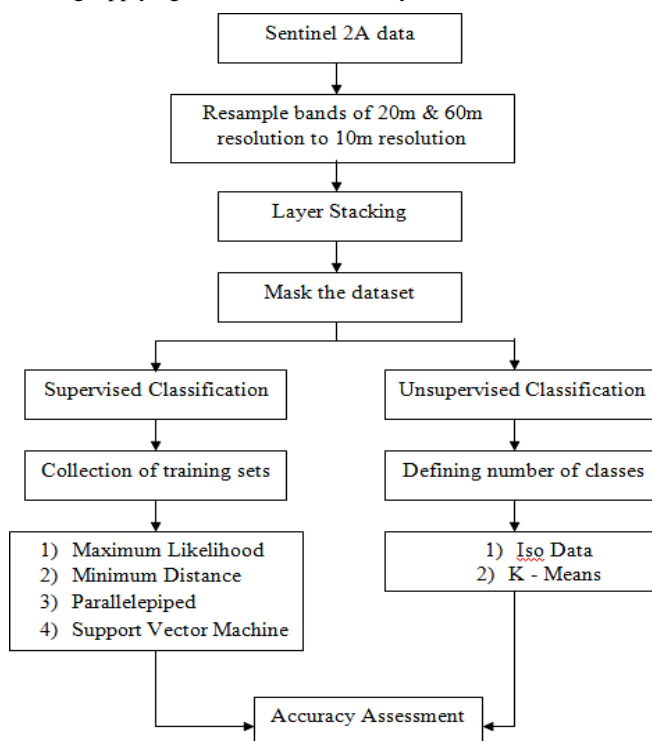


Figure 3. Flowchart

3.2 Resampling

In general, the original remote sensor data consists of spectral bands with varied spatial resolution. The spatial resolution of each band determines the minimal value of the GSD (Ground Sampling Distance) which in turn increases the accuracy. The

GSD is the smallest distance that the sensor covers on the ground. Minimum the distance maximum will be the accuracy of the data. Since Sentinel 2A data consists of 13 bands with varied spatial resolution, resampling is done in order to match similar resolution for all the MSI bands. Sentinel 2A band resolution is mentioned in the above Table 2. From the data, bands 1, 9, 10 (of 60 m resolution) and bands 5, 6, 7, 8A, 11, 12 (of 20 m resolution) are resampled to band 2, 3, and 4 (of 10 m resolution).

3.3 Layer Stacking

Followed by the process of resampling where the resolution of all the 13 bands was matched to 10 m resolution, layer stacking is done. Layer stacking in image processing is the process of combining image derivatives with same spatial resolution. Also, the spatial resolution of 10 m denotes the GSD in ground which is precise and suitable to figure out exact land features present apparently. This provides an accurate output imagery which can be used for further classification.

3.4 Masking

The layer stacked dataset is a raster and it was masked using a polygon feature. The polygon feature of Paonta Sahib was used to extract the raster dataset which is used for classification.

3.5 Classification algorithms

The following were the various types of algorithms that were used to classify the raster data.

I. Supervised Classification

Remote Sensing data used for traditional methods of image classification are used in a wide range of applications now a day (Perumal et al., 2010). Collection of training data called 'representative training samples' for each classes from the remote sensor imagery (false colour composite) is used to acquire results through supervised classification algorithms (Liu). The training samples are the areas of known identity that are also termed as the Region Of Interest (ROI).

- **Maximum Likelihood Classifier** – in which highest probability is assigned to the vector of a class among all the other probabilities of vectors assigned to numerous class. The pixels are assigned to each class based on the threshold value given by the user. If the class probability value is lower than the threshold value set by the user, then the pixels are unclassified (Ahmad et al., 2012).
- **Minimum Distance Classifier** – in which classes that are close to each other are grouped as a unique prototype. It is used to calculate the mean vector for each class and assign pixel to the closest class. (Bhattacharya).
- **Parallelepiped Classifier** – which means a standard threshold, is given in order to check if a particular pixel belongs to the respective class or not. The parallelepiped is based upon the standard deviation (a class limit dimension) threshold from the mean of each classes. If the threshold range is between defined

low and high, the pixels are classified. Areas that do not fall under the specified threshold range are determined to be unclassified.

- **SVM** - which in general applied for a complex dataset where the input numeric attributes are normalized and pre-processed before classifying. It is a non-parametric statistical learning method (Ustuner et al., 2015). Various kernel types such as, linear, polynomial, sigmoid, radial basis function can be set. The kernel type used in this study was the linear SVM Classifier.

II. Unsupervised classification

The unsupervised classification does not require the training samples instead the number of required classes and number of iterations need to be specified.

- **ISO Data** – ISO Data unsupervised classification clusters the pixels that are evenly distributed and groups the remaining pixels based upon defined threshold. The number of classes and iterations need to be given manually.
- **K – Means** - defines the clusters based on the centre pixel of the cluster or assigning each data point as a cluster. The homogenous pixels are clustered together as an object.

3.6 Accuracy Assessment

Using the applied classification results, accuracy assessment is estimated and compared. The accuracy assessment is determined by using the ground truth Region of Interest collected.

4. RESULTS AND DISCUSSION

4.1 Classification Results

The above mentioned supervised and unsupervised classification was applied to the Sentinel 2A data, and the following result shown in Figure 4 was obtained. The data was classified into 7 classes namely, River, Forest, Urban, Mountain, Scrub land, Crop land and River associated sand. For defining the training sets in supervised classification, a maximum number of 30 samples were collected from each class. More the number of classes better would be the aggregation of pixels and hence, number of samples defined was maximum. The size of the samples collected was smaller ellipses and were collected in a distributed manner by covering all the areas the feature was present.

The classification results of Maximum Likelihood, Parallelepiped, SVM provided better results and the LU/LC features were well identified. The results showed maximum similarity with that of the original remote sensor imagery. While the classified images also has misclassified pixels especially in Minimum Distance, ISO data and K – Means due to similarity in the pixels of two varied classes. It is seen that, the shadow region of the hilly mountainous terrain were misclassified as river feature, scrub land misclassified as crop land thus yielding fair accuracy.

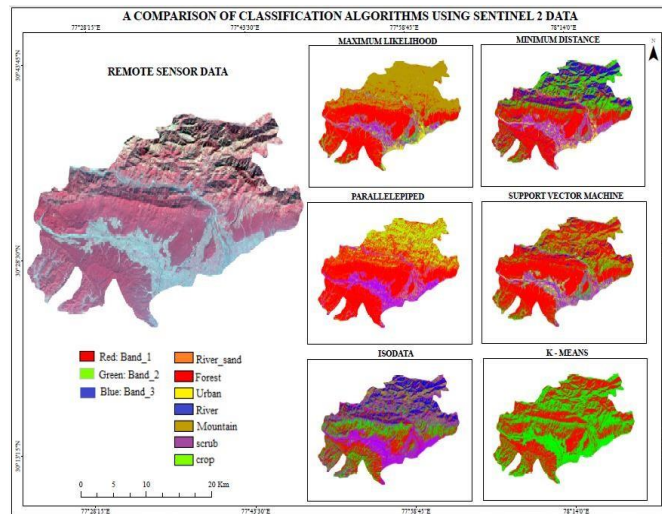


Figure 4. Classification results - comparison

4.2 Estimated accuracy

A confusion matrix displays the reference class and the classified data using which the overall accuracy and kappa coefficient is estimated and compared. The overall accuracy is calculated as a ratio by adding the total number of correctly classified sites to that of the total number of reference sites. This provides a percentage value which is the overall accuracy. The kappa coefficient is a value to evaluate the obtained overall accuracy. In general, kappa coefficient value ranges between -1 to 1 where, -1 denotes poor classification and 1 denotes good classification. The following list of table displays the overall accuracy and Kappa coefficient for each classifier.

Overall								
Accuracy = (464036/519666) 89.30%								
Kappa								
Coefficient = 0.8481								
Ground Truth (Pixels)								
Class	River_sand	Forest	urban	river	Mountain	scrub	crop	Total
Unclassified	0	0	0	0	0	0	0	0
River_sand	56983	0	674	0	271	1804	0	59732
Forest	1	227344	0	110	4569	8	160	232192
urban	12466	30	4334	38	257	7267	30	24422
river [Blue]	20	14	0	8837	18	6	0	8895
Mountain	406	4184	36	23	123314	748	86	128797
scrub	12840	18	928	0	997	37302	58	52143
crop	82	6154	52	4	637	634	5922	13485
Total	82798	237744	6024	9012	130063	47769	6256	519666

Table 5. Maximum Likelihood

Overall								
Accuracy= (321681/519666) 61.90%								
Kappa								
Coefficient = 0.5031								
Ground Truth (Pixels)								
Class	River_sand	Forest	urban	river	Mountain	scrub	crop	Total
Unclassified	0	0	0	0	0	0	0	0
River_sand	59381	11	1048	0	3498	4371	0	68309
Forest	22	200377	3	11	17685	51	342	218491
urban	13058	71	3878	36	233	11120	7	28403
river	41	3594	22	8961	30560	61	0	43239
Mountain	148	1069	35	0	13077	686	990	16005
scrub	10040	135	1025	4	2054	31256	166	44680
crop	108	32487	13	0	62956	224	4751	100539
Total	82798	237744	6024	9012	130063	47769	6256	519666

Table 6. Minimum Distance

Overall								
Accuracy= (416086/519666) 80.07%								
Kappa								
Coefficient = 0.7054								
Ground Truth (Pixels)								
Class	River_sand	Forest	urban	river	Mountain	scrub	crop	Total
Unclassified	656	0	3	0	305	0	0	964
River_sand	81068	240	5876	4	5307	46620	395	139510
Forest	28	231672	1	242	27602	71	979	260595
urban	253	414	135	74	2133	861	701	4571
river	12	14	2	8519	24	16	0	8587
Mountain	781	5404	7	173	94692	201	4181	105439
scrub	0	0	0	0	0	0	0	0
crop	0	0	0	0	0	0	0	0
Total	82798	237744	6024	9012	130063	47769	6256	519666

Table 7. Parallelepiped

Overall								
Accuracy=		(392482/519666)		75.58%				
Kappa		Coefficient						
		0.6546						
Ground Truth (Pixels)								
Class	River_sand	Forest	urban	river	Mountain	scrub	crop	Total
Unclassified	0	0	0	0	0	0	0	0
River_sand	215480	66	23	1	60	41408	160	257198
Forest	23	31913	10961	3809	2	705	87	47500
urban	24	13281	70718	1642	0	1323	18	87006
river	23	1098	668	437	36	597	3	2862
Mountain	97	30	24	10	8893	11289	0	20343
scrub	1700	68	169	5	0	59142	89	61173
crop	20397	1168	174	73	0	15501	5899	43212
Total	237744	47624	82737	5977	8991	129965	6256	519666

Table 8. SVM

Overall								
Accuracy=		(145669/519666)		30.03%				
Kappa		Coefficient						
		0.0917						
Ground Truth (Pixels)								
Class	River_sand	Forest	urban	river	Mountain	scrub	crop	Total
Unclassified	0	0	0	0	0	0	0	0
class 1	0	0	0	0	0	0	0	0
class 2	2458	107722	2345	9012	44773	705	87	47500
class 3	1910	3719	2	0	37277	1323	18	87006
class 4	0	0	0	0	0	3541	3	2862
class 5	0	0	0	0	0	11289	0	20343
class 6	78269	126291	3676	0	38326	59142	3033	61173
class 7	161	12	1	0	9727	15501	5899	43212
Total	82798	237744	6024	9012	130063	5082	48823	519666

Table 9. ISO Data

Overall								
Accuracy=		(93952/519666)		22.09%				
Kappa		Coefficient						
		0.0917						
Ground Truth (Pixels)								
Class	River_sand	Forest	urban	river	Mountain	scrub	crop	Total
Unclassified	0	0	0	0	0	0	0	0
class 1	0	0	0	0	0	0	0	0
class 2	0	0	0	0	0	0	0	0
class 3	0	0	0	0	0	25	18	43
class 4	166	132878	286	9008	45116	3541	1323	182858
class 5	82632	104866	5738	4	84944	11289	0	20343
class 6	0	0	0	0	3	0	6482	6485
class 7	0	0	0	0	0	15501	5899	43212
Total	82798	237744	6024	9012	130063	5082	48823	519666

Table 10. K – Means

5. CONCLUSION

The LU/LC mapping is the most essential layer in various spatial applications. Hence, it needs to be precise in order to proceed with further studies. In spite of the conventional methods of manual digitization, classification provides precise

results along with estimated accuracies. In this classification, the LU/LC features are well interpreted using Maximum Likelihood classifier which provided an overall accuracy of 89.30% followed by Parallelepiped classifier of 80.07%, SVM classifier of 75.58%, Minimum Distance classifier of 61.90%, ISO Data classifier of 30.03% and K – Means classifier of 22.09%. Thus, the Maximum Likelihood Classifier of Supervised classification is the best suited algorithm for classification of LU/LC features.

REFERENCES

- Ahmad, A., & Quegan, S. (2012). Analysis of Maximum Likelihood Classification on Multispectral Data. *Applied Mathematical Sciences*, Vol. 6 (129), 6425 - 6436.
- Al-Ahmadi, F. S., & Hames, A. S. (2008). Comparison of Four Classification Methods to Extract Land Use and Land Cover from Raw Satellite Images for Some Remote Arid Areas, Kingdom of Saudi Arabia. *JKAU; Earth Sci.*, 20 (1), 167-191.
- Amna, B., Shabbir, R., Ahmad, S. S., & Aziz, N. (2015). Land use change mapping and analysis using Remote Sensing and GIS: A case study of Simly watershed, Islamabad, Pakistan. *The Egyptian Journal of Remote Sensing and Space Science*, 18 (2), 251-259.
- Bhattacharya, A. *Introduction To Image Classification*.
- Congalton, R. G. (1991). A review of assessing the accuracy of classifications of remotely sensed data. *Remote Sensing of Environment*, 37(1), 35-46.
- Huang, C., Davis, L., & Townshend, J. (2010). An assessment of support vector machines for land cover classification. *International Journal of Remote Sensing*, 725-749.
- Immitzer, M., Vuolo, F., & Atzberger, C. (2016). First Experience with Sentinel-2 Data for Crop and Tree Species Classifications in Central Europe. *Remote Sens.*, 8(3), 166.
- Kaplan, G., & Avdan, U. (2017). MAPPING AND MONITORING WETLANDS USING SENTINEL IMAGERY. *ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, 271-277.
- Kopecká, M., Szatmári, D., & Rosina, K. (2017). Analysis of Urban Green Spaces Based on Sentinel-2A: Case Studies from Slovakia. *Land*, 6(25), 17.
- Lefebvre, A., Sannier, C., & Corpetti, T. (2016). Monitoring Urban Areas with Sentinel-2A Data: Application to the Update of the Copernicus High Resolution Layer Imperviousness Degree. *Remote Sens.*, 8(7), 606.
- Liu, X. *Supervised Classification and Unsupervised Classification*.
- Madhu, T., Kumar, D. N., Reddy, D. N., Teja, P. V., Narayana, L., Vishal, V., et al. (2017). Spatial-Temporal Analysis of LU/LC Classification in Nirmal Mandal, Adilabad, Telangana

State, India, by Using Remote Sensing and GIS. *International Journal of Geosciences* , 1315-1331.

Melesse, A. M., Weng, Q., Thenkabail, P. S., & Senay, G. B. (2007). Remote Sensing Sensors and Applications in Environmental Resources Mapping and Modelling. *MDPI* , 7 (12), 3209–3241.

Nitheshnirmal, S., Rahaman, S., Balasundareswaran A., (2017). Urban Change Detection and Site Suitability Analysis in Thiruverumbur Taluk, Tiruchirappalli District – A Geospatial Approach. 38th Asian Conference on Remote Sensing - Space Applications: Touching Human Lives, New Delhi.

Perumal, K., & Bhaskaran, R. (2010). Supervised Classification Performance of Multispectral Images. *Journal of Computing* , Volume 2 (2).

Rahaman, S. A., Aruchamy, S., Balasubramani, K., and Jegankumar, R.: LAND USE/LAND COVER CHANGES IN SEMI-ARID MOUNTAIN LANDSCAPE IN SOUTHERN INDIA: A GEOINFORMATICS BASED MARKOV CHAIN APPROACH, *Int. Arch. Photogramm. Remote Sens. Spatial Inf. Sci.*, XLII-1/W1, 231-237, <https://doi.org/10.5194/isprs-archives-XLII-1-W1-231-2017>, 2017.

Rwanga, S. S., & Ndambuki, J. (2017). Accuracy Assessment of Land Use/Land Cover Classification Using Remote Sensing and GIS. *International Journal of Geosciences*, 8, 611-622.

Sekertekin, A., Marangoz, A., & Akcin, H. (2017). PIXEL-BASED CLASSIFICATION ANALYSIS OF LAND USE LAND COVER USING SENTINEL-2 AND LANDSAT-8 DATA. *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences, Volume XLII-4/W6*, pp. 91-93.

Topaloğlu, R. H., Sertel, E., & Musaoğlu, N. (2016). Assessment of Classification Accuracies of Sentinel-2 and Landsat-8 Data For Land Cover / Use Mapping. *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences* , XLI-B8, 12-19.

Ustuner, M., Sanli, F. B., & Dixon, B. (2015). Application of Support Vector Machines for Landuse Classification Using High-Resolution RapidEye Images: A Sensitivity Analysis. *European Journal of Remote Sensing* (48), 403-422.