APPLICATION OF SUPPORT VECTOR MACHINES FOR FODDER CROP ASSESSMENT

Sumi Kala[1*], Magan Singh[1], Sujay Dutta[2], Narendra Singh[3], Shashank Dwivedi[1]

[1]ICAR-National Dairy Research Institute, Karnal-132001, Haryana, India (sumikala, magansingh07, shashank.dwivedi78) @gmail.com
[2]Space Application Centre, ISRO, Jodhpur Tekra, Ambawadi Vistar, Ahmedabad-380015, India (dutta.sujay10@gmail.com)
[3]Indian Institute of Technology (Indian School of Mines), Department of Mining Engineering, Dhanbad-826004, Jharkhand, India (narendrasngh38@gmail.com)

Commission V, SS: Emerging Trends in Remote Sensing

KEY WORDS: Fodder, SVM, MLC, Accuracy, Spectral-temporal, NDVI, Landsat-8

ABSTRACT

Identification of crop and its accuracy is an important aspect in predicting crop production using Remote Sensing technology. This study investigates the ability of Support Vector Machine (SVM) algorithm in discriminating fodder crops and estimating its area using moderate resolution multi-temporal Landsat-8 OLI data. SVM is a non-parametric statistical learning method and its accuracy is dependent on the parameters and the kernels used. The objective was to evaluate the feasibility of SVM in fodder classification and compare the results with traditional parametric Maximum Likelihood Classification (MLC). Fodder crops are available over small fields in the study area thus having large number of pure fodder pixels over small area is difficult. Hence, SVM has an advantage over MLC as it works well with less training data sets also. Three kernels (linear, polynomial and radial based function) were used with SVM classification. Comparative analysis showed that higher overall accuracy was observed in SVM in comparison to MLC. Temporal change in the spectral properties of the crops derived through Normalized Difference Vegetation Index (NDVI) from multi-temporal Landsat-8 was found to be the most important information that affects accuracy of classification. The classification accuracies for SVM with radial based function, polynomial, linear kernel and MLC were 90.09%, 89.9%, 88.9% and 82.4 % respectively. The result suggested that SVM including three kernels performed significantly better than MLC. India has low livestock productivity due to unavailability of fodder hence this study could help in strengthening the fodder productivity.

1. INTRODUCTION

Fodder crops are plant species that are raised for feeding livestock. The most significant three sources of fodder supply are: crop residues (dry fodder), cultivated fodder (green fodder) and fodder from forests, permanent pastures and grazing lands. Most of the fodder requirement is met by feeding crop residues and grazing land. Major fodder crops grown in India are Sorghum (Jowar), Berseem (Egyptian clover), Lucerne (Alfalfa), Pearl millet (Bajra), Maize (Makka), Oats (Jai) and Cowpea (Lobia). Besides improving livestock productivity, fodder crops have use in improving the soil structure, environment protection from pollution, reclamation of degraded land and several others (Squires, 2016). Sorghum among Kharif Crop and Berseem among Rabi crop occupy more than 50% of the total area cultivated under fodder in India (Kumari & Maiti, 2006).

India has one of the largest livestock population in the world however, its productivity is low compared to other developing countries because of the non-availability of proper feed and fodder (Ministry of Agriculture, 2013). Only 4.9% of total cultivated land is devoted for fodder crops leading to a net deficit of 35.6% of green fodder (IGFRI-ICAR).

Remote Sensing technology provides precise, reliable and well-timed data about crop monitoring and yield, thus helping in improving the agricultural statistics system. Accurate crop production forecast using Remote Sensing technology enables policy makers in taking decisions regarding its storage, distribution, procurement of price, import/export policies.

Several studies have been carried out in past for crop assessment using single and multi-date image classification algorithms concluding that multi-date analysis provides better results. Maximum Likelihood classification (MLC) gives more accurate classification results with medium resolution multi-temporal data in comparison to EVM, NN and SAM (Azar et al., 2016). Considering three to four image acquisitions which are well distributed over the growing season, capture the maximum spectral and temporal variability (Roumenina et al., 2015). MLC using multi-temporal medium resolution satellite imagery has the ability to differentiate cropping types, rotations, and irrigation practices in complex regions outperforming the results of single date imagery (Heller et al., 2012).

Currently, Support Vector Machine (SVM) classifier is being widely used for remotely sensed image classification. SVM a Kernel-based, nonparametric technique with various applications including crop classification (Pal, 2009). Pal and Mather, 2004 observed that SVM outdoes MLC and several other classifiers. Unlike MLC, SVM can work well with a small training data (Pal, 2009) set which is great advantage as collecting pure pixel is difficult especially with moderate resolution data. The selection of model parameters and kernel types (linear, polynomial, radial based function and sigmoid) have an important role on SVMs classification accuracy (Ustuner et al., 2015). Azhar et al., 2016 used MLC to classify forage and 6 other classes including rice, maize, soybean, winter

* Corresponding author

crop and arboriculture-woodland. Devdas et al., 2012 used SVM for crop classification but combined fodder crops with other crops.
The work on Fodder crops classification have not been extensively reported as compared to major food crops especially in India. The major problem, in India, in mapping fodder crops is random and small field size. Dedicated large fields for fodder are less as most of the requirement is met by crop residue or grazing land. This leads to difficulty in identifying large number of pure pixels. Therefore, in this study, SVM has been adopted as it has ability to provide better classification with less training set data.

This study describes the usability of MLC and SVM for fodder crop classification moderate resolution multi-temporal Landsat-8 OLI data. The objectives of this study were to
(i) Study the performance of SVM and MLC for fodder crop classification using multi-date Landsat-8 data
(ii) Estimate area under fodder cultivation
The satellite imageries acquired from October, 2016 to February, 2017 were used to derive atmospherically corrected Normalized Differential Vegetation Index (NDVI). Hadjimitsis et al., 2010 highlighted the importance of considering atmospheric effects when NDVI is used for agricultural applications including the study of crop growth patterns. Fodder production has reached a stagnant state and this study would help in development of fodder so as to sustain the livestock by improving the availability of fodder.

2. STUDY AREA

The present study was carried out in Ambala, Kurukshetra and Yamunanagar districts of Haryana state, India (Figure 1). The study areas lies between the 30° 2' 47''to 30° 28' 5''latitudes and 77° 35' 35'' to 77° 3' 38''longitudes bounded by Himachal Pradesh in North, Uttar Pradesh in east, Punjab in west and other districts of Haryana in South. It covers an area of approximately 4894 square kilometers. Yamuna, Markanda, Tangri and Beghnaand rivers mainly drain the district (Ground Water Control Board, 2009). Groundwater is an important source of irrigation. The climate is generally dry during summer period from Mid-March to End of June, followed by southwest monsoon lasting up to September and cold winters (Late November to Early March). The average normal annual rainfall of these districts is 902 mm which is unevenly distributed over the area. The area is mainly covered by silty loam, light loam and loam soils. Because of the presence of sufficient water resources, fertile soil and warm climate agriculture is an important practice of these districts. Some of the more commonly grown fodder crops during the year in the study area are Sorghum, Berseem, Maize, Jowar, Barley and Oats.

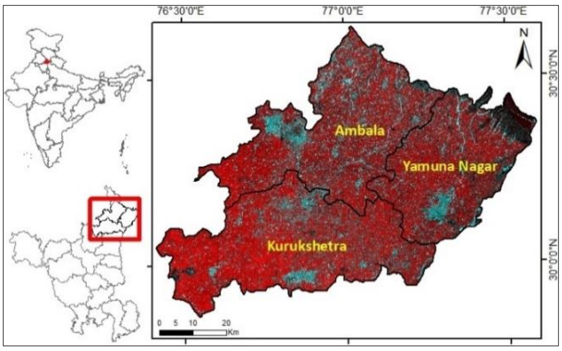

Figure 1. The study area Ambala, Kurukshetra and Yamunanagar districts of Haryana as seen on Landsat-8 OLI standard FCC image of 24-02-2017

3. DATA USED

3.1. Satellite Data

In this study, Landsat-8 OLI data was used for fodder crops assessment during the period October, 2016 to February, 2017. Landsat-8 data was downloaded from USGS Earth Explorer (https://earthexplorer.usgs.gov). Landsat-8 OLI has good repeativity of 16 days, however, the data acquired on 04-11-16, 06-12-16, 07-01-17 and 08-02-17 were highly covered with clouds therefore not taken for further study. Band 1,8,9 have not been used because Band 1 gives ocean color, Band 8 works like a panchromatic film instead of collecting visible colors and Band 9 is used for detecting cirrus contamination in other bands (Ko et al., 2015).

3.2. Ground Truth Data Collection

GPS survey was carried out to randomly collect the location information of fodder crops and other crops in beginning of February 2017 in the study area. Along with the location, other information including field size, crop growth stage, adjacent crop name and a sketch of location was also noted down. These points were used to generate training sites for different classes during classification. Most of the area was covered by Berseem among the fodder crops and wheat & sugarcane among the food crops.

4. METHODOLOGY

4.1. Co-registration

Landsat- 8 image dated 24-02-2017 was considered as reference image and other images were co-registered to this master image. The accuracy of registration was less than 0.25 pixel. Achieving such higher accuracy is important else it would lead to wrong classification for a particular pixel in a set of multi-date image.

4.2. Conversion to Top of Atmospheric (TOA) Reflectance

To reduce the scene to scene variability so as to compare multi-temporal images even from multi-sensors, TOA reflectance values have been used rather than DN/TOA spectral radiance. This is because TOA reflectance (i) eliminates the cosine effect of different solar zenith angles because of the time difference between data acquisitions, (ii) compensates for different values of the exo-atmospheric solar irradiance coming from spectral band differences and (iii) fixes for the variation in the Earth-Sun distance among different data acquisition dates (Chander et al., 2009). DN values were converted to TOA reflectance using reflectance scaling coefficients given in the MTL metadata by the following equation (Landsat 8 Data Users Handbook, 2016):

$$\rho_\lambda' = M_\rho Q_{cal} + A_\rho \quad (1)$$

$$\rho_\lambda = \frac{\rho_\lambda'}{\sin(\theta_{SE})} \quad (2)$$

where,
ρ_λ = Planetary TOA reflectance
ρ_λ' = planetary TOA reflectance, without solar angle correction,
M_ρ =Band-specific multiplicative rescaling factor from the metadata,
A_ρ = Band-specific additive rescaling factor from the metadata,
Q_{cal}= Quantized calibrated digital number (DN) standard product
θ_{SE}= Solar Elevation Angle (from the metadata, or calculated)

4.3. Atmospheric Correction: Histogram Minimum Method

Atmospheric Correction is particularly important for comparison and analysis of multi-temporal images and if the effects of the atmosphere are not considered when deriving vegetation indices from satellite images, may lead to major differences in the final products (Hadjimitsis et al., 2010). Histogram minimum method was applied to multi-date satellite images using the minimum value from each band. These values were subtracted from each TOA reflectance values of their respective spectral band.

4.4. Normalized Differential Vegetation Index (NDVI)

Normalized Difference Vegetation Index (NDVI) is one of the important and widely used vegetation index developed by Rouse et al 1974. NDVI for a pixel varies with different stages of crop growth: at the time of sowing NDVI remains low, rises by maturity and falls again at the time of harvesting. These variations help to discriminate the crops. NDVI of five dates was derived from TOA reflectance images. NDVI values vary for different features; for example, Water & Shadows (NDVI< 0), barren areas & sand (NDVI<0.1), sparse vegetation like grassland, shrubs & senescing crops (0.2≤NDVI≤0.5) and dense vegetation like temperate and tropical forest or crops at their peak stage (0.6 ≤NDVI≤0.9) (Simonetti et al., 2014).

4.5. Signature Generation and Evaluation

NDVI images of five dates were stack to generate the spectral-temporal profiles. Training sites were generated from the ground truth data collected. Several training sites were merged into 7 sites namely Fodder crops, Other Crops (Wheat, Sugarcane and Other Crops), plantation, forest, urban, Sand and water were created. All other classes were merged into one as increasing the number of class reduced accuracy and in this study the focus was to identify fodder crops (Fernandes, 2015). The major problem in generating sites for fodder crops was small field size therefore, at a time not more than 2-3 pixels were taken to define a particular training site in Fodder class which were later on merged together.

4.6 Maximum Likelihood Supervised Classification

Maximum likelihood supervised classification algorithm has been used for over the years worldwide for crop classification. MLC is a parametric classifier taking into account the variance covariance within the class distributions thus for normally distributed data it performs better than the other known parametric classification techniques. However, most of the time the data is not normally distributed hence the results could be unsatisfactory.

4.7 Support Vector Machine

SVM is a linear binary classifiers that assign a given pixel a class from one of two possible classes by separating the data through a hyper plane. SVM has several advantage in image classification because (i) SVM show good performance with a small training data set as it is difficult to select sufficient number of pure training pixels, (ii) SVM good accuracy for datasets having large number of features (Gualtieri and Cromp, 1998) and (iii) SVM are robust to the over fitting problem (Pal, 2012). The accuracy of SVM is also dependent on the choice of kernel. In this study we have used three kernels e.g. linear, polynomial and radial based function (RBF). The following parameters were identified on SVMs classification (i) error penalty or cost (C = 100) for the all kernels, (ii) gamma (γ =0.2) for all kernel types except linear, (iii) bias term (r=1) for polynomial (iv) polynomial degree (d=2) for polynomial kernel.

4.8 Accuracy Assessment

Accuracy assessment is done by generating error matrix or confusion matrix which describes the relationship between the two sources of information i.e. (i) pixels derived from classification map, and (ii) ground reference test information obtained at the same location. Kappa coefficient is used in this study to reflect the measure of difference in actual agreement and expected agreement. Ground reference points used for accuracy assessment were different from the ones used in training set generation to avoid biasness.

4.9 Area Estimation

Area estimation was carried out through detailed enumeration by pixel count method. Pixel count is taken from the histogram of classified image. Each pixel represents an area of 30*30=900 square meters.

$$\text{Area(hectares)} = \frac{(\text{Pixel Count})*(30*30)}{10000} \qquad (3)$$

5. RESULTS

The present study investigates the potential of classification algorithms in discriminating fodder crops. NDVI derived from multi-temporal Landsat-8 images was classified using three SVM kernels (linear, polynomial and radial based function). The MLC was applied on the same dataset for comparative analysis. Both algorithms showed that the spectral temporal response based on NDVI is an important variable which demonstrates the different phonological stages like time of sowing, green-up, maximum greenness and senescence. Training set was derived for different class from the ground reference data. Classes including urban area, water bodies and river beds were not required hence masked out for the analysis. A brief description of the main classes identified (Figure 2) is as follows:

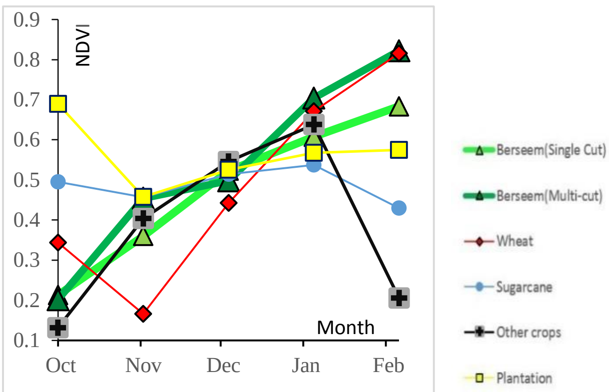

Figure 2. NDVI based temporal profiles for different vegetation classes

(i) Fodder Crops: Berseem is the major fodder crop cultivated in the study area, sown by mid to end of October. Single cut Berseem shows a continuous increase in NDVI with time. Multi-cut Berseem shows dip in NDVI at end of December and rises thereafter.

(ii) Other Crops: It includes all other crops mainly including wheat, sugarcane. Wheat has maximum cultivated area in this region during Rabi season. It is mostly sown by mid-end of November. In January & February NDVI values of wheat approach very close to that of Berseem hence it is quite difficult to distinguish them based on their maturity. The difference in the date of sowing of Berseem and Wheat was considered as an important factor in distinguishing both of them.
(iii) Plantation: Since plantation remains dominant round the year, this class maintained an average NDVI >=0.5 for the entire season.
(iv) Forest: A small north east part of study area is covered by forest which also shows high NDVI values for most of the season.
Tonal variations of Fodder crop corresponding to spectral growth profile is shown in Figure 3.

Figure 3. The tonal variation in reflectance for Berseem and Wheat

Accuracy assessment was carried out for four classes were considered i.e. Fodder Crops, Other Crops, Plantation and Forests. Accuracy assessment distinctly indicated that SVM (Figure 4) and MLC (Figure 5) both performed well (OA >80%) however, SVM showed higher potential for fodder crop classification. SVM also provided speckle free image which is important as speckle effect leads to overestimation of fodder crops in MLC (Figure 6). The reason for MLC giving less accuracy then SVM could be because of the training set does not have sufficiently large number of pure pixels Amongst the three kernel used, radial based function and polynomial showed equally good accuracies. However, SVM kernels can outperform any other kernel because the performance of kernel can vary with the data and training set (Pal, 2012). Also, changing the kernel parameter (C, gamma, d) can vary the accuracy of classification (Ustuner et al., 2015). The SVM model with RBF kernel generated classified image with an overall accuracy of 90.09% (k=0.85) (Table 1) whereas MLC had an overall accuracy of 82.4% (k=0.74) (Table 2). Incorrect classification of Fodder crops at some places in both SVM and MLC could be because of the any of the following reasons (i) Late sown Berseem and Early sown Wheat (beginning of November) will show similar trend (ii) Small size of Fodder fields limiting the identification.

The total area occupied by fodder crops was calculated by direct pixel based counting method (Figure 7). Because of the overestimation in MLC final area calculations are given based on SVM. The total area under fodder cultivation in three district was estimated as 6.39 per thousand hectares.

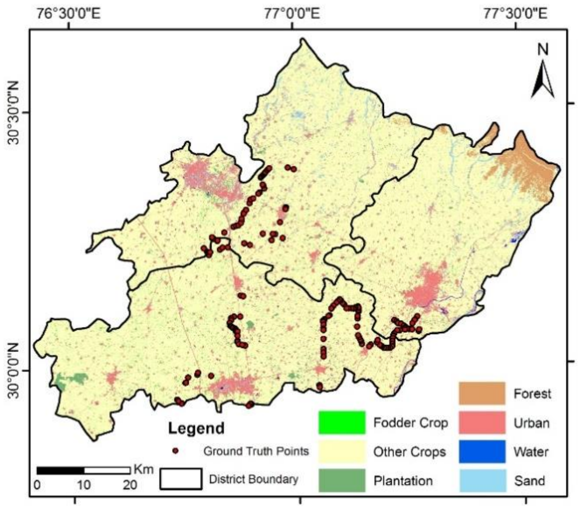

Figure 4. SVM (RBF) image of Study Area with ground truth points.

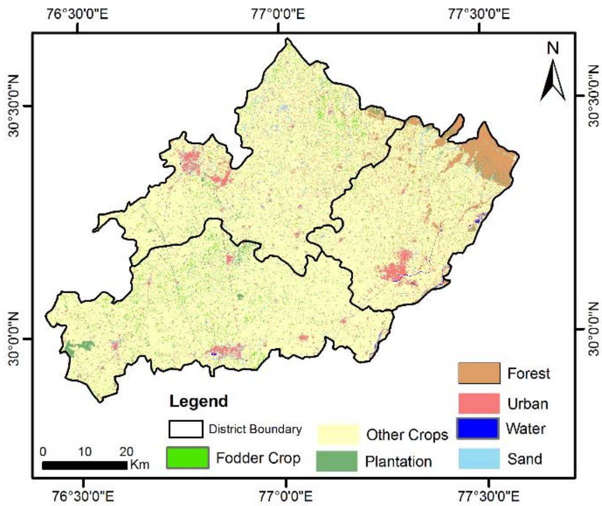

Figure 5. MLC image of study area with Ground truth Points

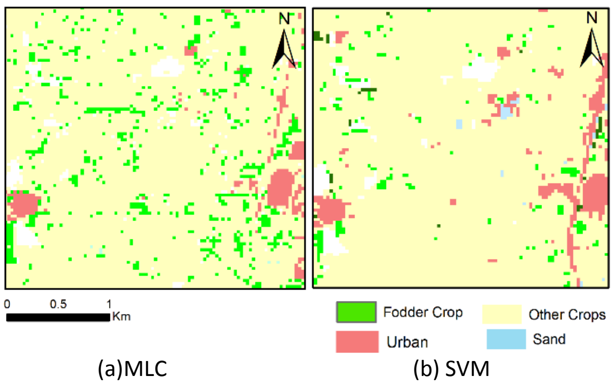

Figure 6. Speckle effect is significantly higher in MLC as compared to SVM for the same area.

Classified Data	Reference Data				Classified Totals
	Fodder Crops	Other Crops	Plantation	Forest	
Fodder Crops	28	5	0	0	33
Other Crops	2	90	10	0	102
Plantation	0	3	51	1	55
Forest	0	0	1	31	32
Reference Total	30	98	62	32	222

Table 1: Confusion Matrix for SVM (RBF-kernel)

Classified Data	Reference Data				Classified Totals
	Fodder Crops	Other Crops	Plantation	Forest	
Fodder Crops	23	13	0	0	36
Other Crops	7	78	8	0	93
Plantation	0	7	50	0	57
Forest	0	0	4	32	36
Reference Total	30	98	62	32	222

Table 2: Confusion Matrix for MLC

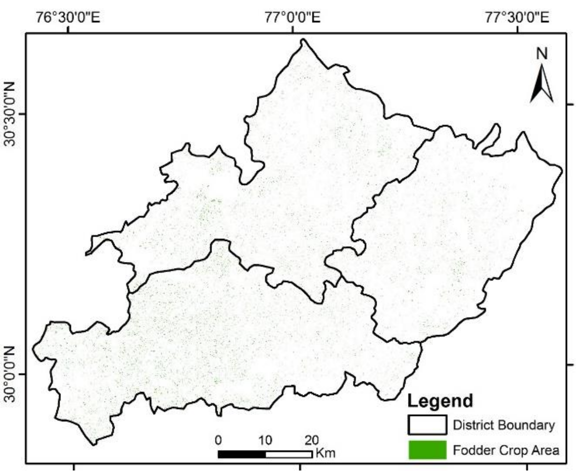

Figure 7. Area under Fodder Cultivation

6. CONCLUSION

An attempt was made to analyze the performance of SVM and MLC in discriminating the Fodder Crops using Landsat-8 OLI imagery for Rabi season. Remote Sensing technology helps to map the sparsely distributed fodder crops in the study area. NDVI based spectral temporal characteristics were basis of classification algorithms. Three SVM kernels were considered for analysis and all of them showed higher accuracy than MLC. SVM had an overall accuracy of 90.09 % for RBF and 89.99% for polynomial which is significantly greater than MLC with 82.4 %.

Fodder crop area was estimated from both the classified maps and it was observed that MLC did overestimation of fodder area. Kurukshetra showed the highest area whereas lowest area was in Yamunanagar as it is mostly devoted to sugarcane cultivation. Results indicate that SVM has greater potential in discriminating fodder crops with a small training datasets in comparison to MLC using moderate resolution data. Such study can be further carried out to have the fodder statistics for the entire country throughout the year.

ACKNOWLEDGEMENT

Authors are grateful to Director, Space Applications Centre, Ahmedabad and Director, ICAR-NDRI, Karnal for their encouragement and support for carrying out this work.

REFERENCES

Azar, R., Villa, P., Stroppiana, D., Crema, A., Boschetti, M., Brivio, P.A., 2016. Assessing in season crop classification performance using satellite data: A test case in Northern Italy. Eur. J. Remote Sens. 49, pp. 361–380. doi:10.5721/EuJRS20164920

Chander, G., Markham, B.L., Helder, D.L., 2009. Summary of Current Radiometric Calibration Coefficients for Landsat MSS, TM, ETM+, and EO-1 ALI Sensors. Remote Sens. Environ. 1, pp. 1–24.

Department of Animal Husbnadry Dairying and Fisheries, Mininstry of Agriculture, 2013. National Livestock Policy.

Devadas, R., Denham, R.J., Pringle, M., Centre, R.S., Percinct, E., 2012. Support Vector Machine Classification of Object-Based Data for Crop Mapping , Using Multi-Temporal Landsat Imagery. Int. Arch. Photogramm. Remote Sens. Spat. Inf. Sci. XXXIX, pp. 185–190. doi:10.5194/isprsarchives-XXXIX-B7-185-2012.

Luís, J., Fernandes, R., 2015. Analysis of Classification Algorithms for Crop Detection using LANDSAT 8 images.

Ground, C., Board, W., 2009. Ground Water Information Booklet.

Gualtieri, J. A. and Cromp, R. F., 1998. Support vector machines for hyperspectral remote sensing classification. Proceedings of the of the SPIE, 27th AIPR Workshop: Advances in Computer Assisted Recognition, Washington, DC, October 14-16, pp. 221-232.

Hadjimitsis, D.G., Papadavid, G., Agapiou, A., Themistocleous, K., Hadjimitsis, M.G., Retalis, a., Michaelides, S., Chrysoulakis, N., Toulios, L., Clayton, C.R.I., 2010. Atmospheric correction for satellite remotely sensed data intended for agricultural applications: impact on vegetation indices. Nat. Hazards Earth Syst. Sci. 10, pp. 89–95. doi:10.5194/nhess-10-89-2010.

Heller, E., Rhemtulla, J.M., Lele, S., Kalacska, M., 2012. Mapping crop types, irrigated areas, and cropping intensities in heterogeneous landscapes of southern India using imagery: Implications for assessing water use in agriculture. Photogramm. Eng. Remote Sens. 78, pp. 815–827. doi:10.1117/12.697882.

ICAR-IGFRI, 2013. Vision 2050.

Interior, D. of the, Survey, U.S.G., 2016. LANDSAT 8 (L8) DATA USERS HANDBOOK.

Ko, B.C., Kim, H.H., Nam, J.Y., 2015. Classification of Potential Water Bodies Using Landsat 8 OLI and a Combination of Two Boosted Random Forest Classifiers. Sensors 15, pp. 13763–13777. doi:10.3390/s150613763.

Kumari, A., Maiti, R., 2016. Climate Change: Its Impact on Bio-resources and Sustainable Agriculture, in: Ratikanta, M., Kumari, A., Thakur, A.K., Sarkar, N.C. (Eds.), Bioresources and Stress Management. Springer, pp. 13–32.

Pal, M., 2009. Kernel methods in remote sensing: A review. ISH J. Hydraul. Eng. 15, pp. 194–215. doi:10.1080/09715010.2009.10514975.

Pal, M., 2012. Advanced algorithms for land use and cover classification. Advances in Mapping from Remote Sensor Imagery, CRC Press, pp. 69-90. doi: http://dx.doi.org/10.1201/b13770-4.

Pal, M. and Mather, P. M. 2004. Assessment of the effectiveness of support vector machines for hyperspectral data. Future Generation Computer System, 20: pp. 1215–1225.

Roumenina, E., Atzberger, C., Vassilev, V., Dimitrov, P., Kamenova, I., Banov, M., Filchev, L., Jelev, G., 2015. Single- and multi-date crop identification using PROBA-V 100 and 300 m S1 products on Zlatia Test Site, Bulgaria. Remote Sens. 7, pp. 13843–13862. doi:10.3390/rs71013843.

Rouse, W., Haas, R.H., Deering, D.W., 1974. Monitoring vegetation systems in the Great Plains with ERTS. Proceedings of the Third Earth Resources Technology Satellite- 1 Symposium, pp. 301–317.

Simonetti, E., Simonetti, D., Preatoni, D., 2014. Phenology-based land cover classification using Landsat 8 time series. doi:10.2788/15561.

Squires, V.,2011. The role of food, agriculture, forestry and fisheries in human nutrition. Vol. 1 Forage crop production-The role of food, agriculture, forestry and fisheries in human nutrition. Adelaide: Eolss.

Ustuner, M., Sanli, F.B., Dixon, B., 2015. Application of support vector machines for land use classification using high-resolution rapid eye images: A sensitivity analysis. Eur. J. Remote Sens. 48, pp. 403–422. doi:10.5721/EuJRS20154823.