https://doi.org/10.56093/ijas.v90i3.101473
Indian Journal of Agricultural Sciences **90** (3): 551–5, March 2020/Article

Remote Sensing and Machine Learning techniques for acreage estimation of mango (*Mangifera indica*)

ASHIS RANJAN UDGATA, PRACHI MISRA SAHOO*, TAUQUEER AHMAD, ANIL RAI and GOPAL KRISHNA

ICAR-Indian Agricultural Statistics Research Institute, Pusa, New Delhi 110 012, India

Received: 22 March 2019; Accepted: 02 August 2019

ABSTRACT

Mango (*Mangifera indica*) is the most important commercially grown fruit crop in India. Uttar Pradesh, Andhra Pradesh, Karnataka, Bihar, Gujarat and Tamil Nadu are the major producers of mango. It covers around 42% total area accounting for 40% of total production in the country. Hence, development of reliable and timely estimates of area under mango at national level is essential for policymakers and planners for market planning and export. Earlier only survey technologies were used for area estimation which was a time consuming and laborious process. Modern space technology like remote sensing can be used as an alternative. Therefore, a study was carried out for acreage estimation of mango in West Godavari district of Andhra Pradesh using Sentinel 2 satellite data in the year 2017. Acreage estimation of mango was done after the preparation of land use and land cover map. Three supervised classification techniques, viz. Maximum Likelihood Classification (MLC), Support Vector Machine (SVM) and Artificial Neural Network (ANN) were used for land use and land cover map preparation. Support Vector Machine using three different kernel functions, viz. Radial Basis Function (RBF), Sigmoid kernel and Polynomial kernel were used to improve the classification accuracy. SVMRBF was found to be the best classification technique with overall accuracy of 94.44 and kappa coefficient 0.9218. The mango area obtained from the classified satellite image using SVMRBF was 9372.96 ha.

Key words: Artificial Neural network (ANN), Maximum Likelihood Classifier (MLC), Sentinel and Support Vector Machine (SVM)

Mango (*Mangifera indica*) is the predominant horticultural crop cultivated in 42% of total area and contributes 40% of total production (Economic Survey of India 2015–16) in India. Although, India has emerged as a leading producer of horticultural crops in the world, official estimates of area under cultivation are not available for mango. This is one of the major data gap in our Agricultural Statistical System. In modern era, space technologies like Remote Sensing (RS) may be explored for quick area estimation. Researchers have used remote sensing for mapping of Leaf Area Index (LAI) (Tripathi *et al.* 2013) and wheat grain yield estimation (Pradhan *et al.* 2013).

Remotely sensed images commonly involve use of the reflectance and radiances of each pixel to assign it to a number of land cover classes (Huang *et al.* 2001). Numerous classification techniques are explained in literature for preparation of land use and land cover map. Maximum Likelihood Classification (MLC), Support Vector Machine (SVM) and Artificial Neural Network (ANN) are the prominent one. The MLC is well-known supervised classification, which minimizes Bayesian error (Duda and Hart 1973). MLC method was used for area estimation of cereals (Patel *et al.* 1991), mango (Yadav *et al.* 2002) and coffee (Rao *et al.* 2002). A methodology for acreage estimation under agroforestry using LISS-IV data in Ludhiana district of Punjab State has been developed from ICAR-IASRI (Ahmad *et al.* 2015). Nowadays SVMs are used in various applications such as text description (Joachims T 1998), face recognition and image classification (Huang *et al.* 2002). Support Vector Machines (SVM) can be used for satellite image classification (Pal 2008, Yang *et al.* 2011). ANN was used for land use and land cover classification (Lu and Weng 2007, Jensen *et al.* 2009, Goswami *et al.* 2014). Many researchers have made comparison between classification techniques (Dixon and Candade 2008). Keeping above points in view, a study was undertaken to select best classification technique for delineation of mango in West Godavari district of Andhra Pradesh and mango area estimation by using sentinel 2 satellite data.

MATERIALS AND METHODS

West Godavari district of Andhra Pradesh was taken for acreage estimation of mango in year 2017. It is spread

*Corresponding author e-mail: prachi.iasri@gmail.com

over 16.706041°N to 17.119720°N and 80.991595°E to 81.741431ºE with an approximate area of 7,742 km^2. Administrative headquarter of West Godavari is situated at Eluru. It receives a rainfall of 785.6 mm in South-west monsoon amounting to 68% and 254.6 mm in North East monsoon which amounts to 21% of the total rainfall. Mainly Alluvial, Black Regur and Red Ferruginous soils along with a small portion of sandy soil in coastal belts are found in this district. Paddy, Sugarcane, Banana, Mango, Cashewnut, Coconut and Cocoa are major crops grown in this district.

Sentinel 2 satellite data was acquired for acreage estimation of mango. Sentinel 2 satellite data, developed by European Space Agency (ESA) provides 13 multispectral bands including Visible and Near-Infrared (VNIR) to Shortwave Infrared. It aims at collecting information on forest watching, land cover changes exposure, and natural disaster management. Three sentinel 2 images were acquired on 28th December 2016 along with another image acquired on 28th November 2016 to cover the whole district. It also enabled to differentiate mango crop from other crop because this period coincides with mango flowering period. The details of four images are given in Table 1.

Training data evaluation: Field surveys were conducted in November 2016 to December 2016 using handheld GPS device. GPS coordinates at each site were recorded and all types of land covers in the district were studied. A sufficient number of points had been collected from which 70% points were used for selecting training samples and remaining were used as validation of sample points. As per the demographic features and the characteristics of study area, as well as research objectives, a classification system with eight categories was designed. The eight land categories were mango orchards, water bodies, forest, built up, dense vegetation areas, wetland, and agricultural land. Agricultural lands are the areas excluding mango areas. By considering the data on the abovementioned classes Region of Interest (ROI) was established by choosing one or more polygons for each class.

Image Pre-processing: After acquisition of satellite data; layer stacking, mosaicking, ROI preparation and image classification were done to get classified map of West Godavari district. Layer stacking is a process of forming a new multiband image from individual band images of various pixel size, extent and projection system. It is a process of combining different bands of raw image (single band) one after another so that the input bands can be resampled and re-projected to a common user-selected output projection and pixel size. All the 13 bands of Sentinel-2 were layer stacked to form a single image by nearest neighbor resampling technique. Image was projected in Universal Transverse Mercator co-ordinate system (UTM) with datum World Geodetic System (WGS) 1984.

Mosaicking is meant for joining individual scenes to form a combined image covering the entire study area. Four scenes were mosaicked on the basis of geo-referenced format to obtain the image of entire district. For mosaicking of scenes, the images should be acquired in same year and Residual Mean Square (RMS) should not be more than 2.5%. RMS in this study was 0.0001, hence it was easily mosaicked. Then the output image was made in False Colour Composiatae (FCC) format by considering band 4, band 3 and band 2.

Image classification : Three supervised classification techniques such as Maximum likelihood classification (MLC), Support vector machines (SVMs) and Artificial Neural Network (ANN) were applied with help of ENVI software. MLC is the most popular and reliable supervised classification method in remote sensing (Duda and Hart 1973). In this, classification is done on the basis of inclusion probability of an unknown pixel. The unknown pixels are classified into the class having maximum probability. The probability threshold for each class is user defined. In this study probability threshold was taken as 50.

SVM is a supervised, non-parametric statistical learning technique, not requiring any assumption of the distribution of underlying data (Pal 2008). Kernel functions are useful when the data are not linearly separable. Three kernel functions were used – Polynomial, Radial Basis Function (RBF) and Sigmoid kernel.

The parameters of the RBF kernel function are penalty parameter, pyramid levels. Penalty parameter value controls the training error for SVM. It takes floating point values more than zero. Number of pyramid levels is a dynamic value used in SVM training and classification process. If number of pyramid level is zero the image is processed in full resolution. Here, penalty parameter was taken as 100

Table 1 Scene characteristics of Sentinel 2 data

Acquisition date	Attribute name	Cloud cover	Orbit no.	UTM zone
December 28, 2016	S2A_OPER_MSI_L1C_TL_SGS__20151228T051728_20151228T102532_A002690_T44QND_N02_01_01	0	119	44N
December 28, 2016	S2A_OPER_MSI_L1C_TL_SGS__20151228T051728_20151228T102532_A002690_T44QME_N02_01_01	3	119	44N
December 28, 2016	S2A_OPER_MSI_L1C_TL_SGS__20151228T051728_20151228T102532_A002690_T44QND_N02_01_01	0	119	44N
November 28, 2016	S2A_OPER_MSI_L1C_TL_EPA__20151128T051517_20170430T185220_A002261_T44QNE_N02_04_01	0.253	119	44N

and pyramid levels and classification probability threshold were taken as their default value 0 (Zero).

There are five parameters in polynomial kernel function such as degree of polynomial kernel, bias in kernel function, gamma kernel function, penalty parameter and pyramid levels. Degree of a polynomial kernel function varies from 1 to 6. The minimum value of the gamma parameter is greater than zero. The default value is taken as the inverse of number of input bands used in input image. In polynomial kernel function the degree of 4, bias 1.0, gamma kernel of 0.077, penalty parameter 100 along with zero pyramid level were used in classification.

The parameter of the sigmoid kernel are bias, gamma kernel function, penalty parameter, pyramid level and classification probability threshold. Gamma kernel value, penalty parameter, bias, penalty parameter and classification threshold value were taken as 0.077, 100, 1.0, 100 and 0, respectively.

ANN is a biological neuron model that is a mathematical description of the pro cells. It mimics the properties of human neuron (Saini *et al.* 2015). The layer within input and output layer is called hidden layer. Hidden layer acts as individual feature detector, and it is also capable of recognizing complex patterns. Initially weights are given random values and the network is trained in order to find the weight parameters that produce the desired effect. To achieve this, comparison between the desired output and network output are to be made. In this study, logistic activation function with one hidden layer has been used. Training threshold contribution, training rate, training momentum, training RMS exit criteria and training iteration were fixed at 0.9, 0.2, 0.9, 0.1 and 100, respectively.

Accuracy assessment: Overall accuracy and the kappa coefficient were commonly used for accuracy assessment (Rosenfield G H and Fitzpatrick-Lins K 1986). Thirty percent of ground truth points were preserved for accuracy assessment i.e. for validation purpose. As per this ground truth points again a Region of Interest (ROI) was prepared and accuracy assessment of the classified image was further done. Kappa coefficient is a statistical test for examining the divergence between two algorithms.The value of Kappa coefficient varies between 0 and 1. K>0.80 indicates good accuracy, 0.40<K<0.80 indicates medium accuracy and K<0.40 shows poor accuracy.

RESULTS AND DISCUSSION

There are four stages of land cover classification such as designing a proper classification system, collection of sufficient and well representative reference data, selection of suitable variables and classification algorithm, and evaluation of the classification result (Lu and Weng). In area of remote sensing our main focus should be on selection of classification system and classification algorithm. On the basis of landscape characteristics and availability of remote sensing data it is important to determine the variables that can be used in separation of different land cover types. Landcover classification accuracy cannot be improved by selection of improper combination of variables (Li *et al.* 2012). Another important feature in classification is time of data acquisition, i.e. vegetation types have different spectral

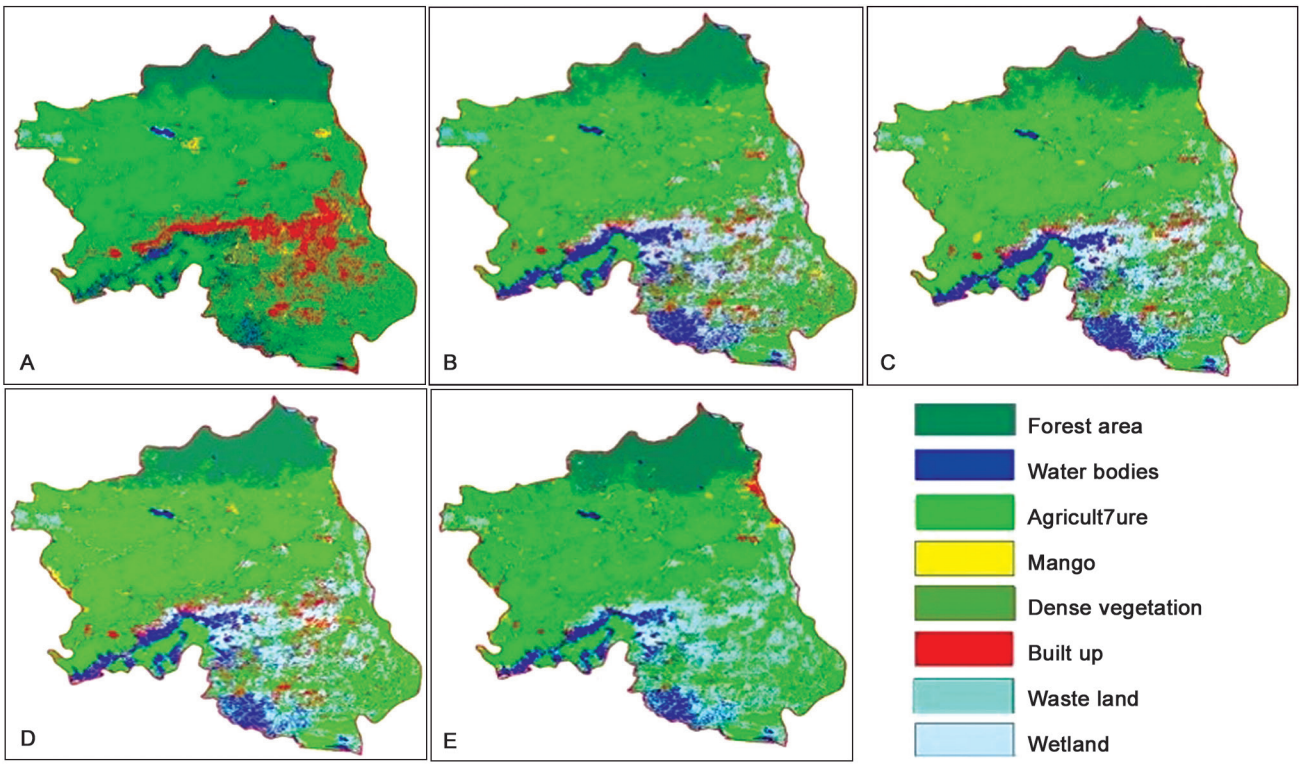

Fig 1 A. Maximum Likelihood Classified image, B. SVMRBF classified image; C. SVM polynomial kernel classified image; D. SVM sigmoid kernel classified image; E. Artificial Neural Network classified image.

Table 2 Accuracy assessment of three classification method

Classification method	Overall accuracy	Kappa coefficient
Maximum Likelihood Classification	79.98	0.7251
SVM Radial Basis Function	94.44	0.9218
SVM Polynomial kernel	93.20	0.8986
SVM Sigmoid kernel	90.61	0.8585
Artificial Neural Network	85.45	0.7945

signatures in different time periods. The accuracy of land cover classification can be improved by using different temporal data. Especially cropland and forest classification can be improved by using medium resolution images (e.g. Landsat) or by coarse resolution images (e.g. MODIS) (Zhu G and Blumberg 2002, Lu and Weng 2007).

The MLC classified image of West Godavari is shown in Fig 1A. The pine green, blue, shamrock green, yellow, juniper green, red, sky blue and snow white color in the MLC classified image shows the forest area, water bodies, agricultural area, mango, dense vegetation areas, built up, waste land and wetland pixels, respectively. From the classified image it is evident that some portion of wetland was classified into built up area. In the classified image 32410 pixels were classified under mango covering an area of 11667.6 ha.

Fig 1B represents the classified image by SVM Radial Basis Function kernel. In all, 26036 number of pixels were identified under mango class in the classified image using SVMRBF covering an area of 9372.96 ha. In the SVM polynomial kernel classified image (Fig 1C) 23285, number of pixels were classified as mango covering 8382.6 ha of area. It was also observed that in SVM sigmoid kernel classified image (Fig 1D) 24836 pixels come under mango covering an area of 8940.6 ha. In Artificial Neural Network classified image (Fig 1E) 25121 number of pixels were identified under mango class covering an area of 9043.56 ha.

Performance of various classifiers was assessed on the basis of overall accuracy and Kappa coefficient (Table 2). Results show that SVMRBF has highest overall accuracy (94.44) followed by SVM polynomial kernel (93.20), SVM Sigmoid kernel (90.61), ANN (85.45) and MLC (79.98). Similar trend was observed in case of kappa coefficient. Highest value of kappa coefficient was obtained by SVMRBF classifier (0.9218). The kappa coefficient value was least for MLC (0.7251). In this research, the three approaches perform differently because of mechanism of classification, choice of data types and time of data acquisition.

From the study, it can be concluded that SVM-RBF has the potential to map land cover distribution based on high spatial resolution satellite images among the classifiers like MLC and ANN. This research had also confirmed the necessity of using SVM-RBF in mango area identification. In future, more research is required to study the combination of different source data in different seasons in purview of land cover classification.

ACKNOWLEDGEMENTS

We acknowledge the help rendered by MNCFC (Mahalanobis National Crop ForecastingCentre), New Delhi for providing ground truth points.

REFERENCES

Ahmad T, Sahoo P M and Jally S K. 2015. Estimation of area under agroforestry using high resolution satellite data. *Agroforestry System* **90**(2): 289–303.

Atkinson P M and Tatnall A R L. 1997. Introduction neural networks in remote sensing. *International Journal of Geographic Information Systems* **18**(4): 699–709.

Dakshinamurti C, Krishnamurthy B, Summanwar A S, Shanta P and Pisharoty P R. 1971. Remote Sensing for Coconut Wilt. *(In) Proceedings of International Symposium on Remote Sensing of Environment,* Ann Arbor, Michigan, May 17-21, pp. 25-9.

Dadhwal V K. 1999. Remote sensing applications for agriculture: retrospective and perspective. *(In) Proceedings of National Symposium on Remote Sensing Applications for Natural Resource and Perspective*, Bengaluru, January 19-21, pp. 11-2.

Dixon B and Candade N. 2008. Multispectral land use classification using neural networks and support vector machines: one or the other, or both. *International Journal of Remote Sensing* **29**(4):1185–1206.

Duda R and Hart P E. 1973. *Pattern Classification and Scene Analysis.* 2nd edn, pp. 45-55. John Wiley & Sons, New York.

GoswamiA K, Gakhar S and Kaur H. 2014. Automatic object recognition from satellite images using artificial neural network. *International Journal of Computer Application* **95**(10): 33–9.

Huang C, Townshend J R G, Liang S, Kalluri S N V and Defries R S. 2001. Impact of sensor's point spread function on land cover characterization: assessment and deconvolution. *Remote Sensing of Environment* **80**(2): 203–12.

Huang C, Davis L S and Townshend J R G. 2002. An assessment of support vector machines for land cover classification. *International Journal of Remote Sensing* **23**(4):725–49.

Jensen R R, Hardin P J and Yu G. 2009. Artificial neural networks and remote sensing. *Geography Compass* **3**(2):630–46.

Joachims T. 1998. Text categorization with support vector machines-learning with many relevant features. *(In) Proceedings of European Conference on Machine Learning,* Chemnitz, Germany, April 21-23, pp. 37–142.

Li G, Lu D, Moran E and Sant'Anna S J S. 2012. A Comparative Analysis of Classification Algorithms and Multiple Sensor Data for Land Use/Land Cover Classification in the Brazilian Amazon. *Journal of Applied Remote Sensing* **6**(1): 0617061-23.

Lu D and Weng Q. 2007. A survey of image classification methods and techniques for improving classification performance. *International Journal of Remote Sensing* **28**(5):823–70.

Pal M. 2008.Ensemble of support vector machines for land cover classification. *International Journal of Remote Sensing* **29**(10): 3043–9.

Pal M and Mather P M. 2005. Support vector machines for classification in remote sensing. *International Journal of Remote Sensing* **26**(5):1007–11.

Patel N K, Ravi N, Navalgund R R, Dash R N, Das K C and Patnaik S. 1991. Estimation of rice yield using IRS 1A digital data in coastal tract of Orissa. *International Journal of Remote Sensing* **12**(11): 2259–66.

Pradhan S, Bandyopadhyay K K, Sahoo R N, Sehgal V K, Singh

R, Joshi D K and Gupta V K. 2013. Prediction of wheat (*Triticum aestivum*) grain and biomass yield under different irrigation and nitrogen management practices using canopy reflectance spectra model. *Indian Journal of Agricultural Sciences* **83**(11): 1136–43.

Rosenfield G H. and Fitzpatrick-Lins K. 1986. A coefficient of agreement as a measure of thematic classification accuracy. *Photogrammetric Engineering & Remote Sensing* **52**(2): 223–7.

Tripathi R, Sahoo, R N, Gupta V K, Sehgal V K and Sahoo P M. 2013. Retrieval of Leaf Area Index using IRS-P6, LISS-III data and validation of MODIS LAI product (MOD15 V5) over trans Gangetic Plains of India. *Indian Journal Agicultural Sciences* **83**(4)**:** 380–5.

Yadav I S, Rao N K S, Reddy B M C, Rawal R D, Srinivasan V R, Sujatha N T, Bhattacharya C, Rao P P N, Ramesh K S and Elango S. 2002. Acreage and production estimation of mango orchards using Indian Remote Sensing (IRS) satellite data. *Scientia Horticulture* **93**(2):105–23.

Yang C, Everitt J H and Murdenb D. 2011. Evaluating high resolution SPOT 5 satellite imagery for crop identification. *Computers and Electronics in Agriculture* **75(**11**)**: 347–54.

Zhu G and Blumberg D G. 2002. Classification using ASTER data and SVM algorithms: The case study of Beer Sheva, Israel. *Remote Sensing of Environment* **80**(2): 233–40.