Quest Journals
Journal of Research in Agriculture and Animal Science
Volume 10 ~ Issue 5 (2023) pp: 07-13
ISSN(Online) : 2321-9459
www.questjournals.org

Research Paper

Accuracy assessment of Principal Component Analysis (PCA) and Supervised Classification method for Crop Acreage Estimation in Dhampur, India.

Dr. Mubashir Jamil[1*]and Vijay Lakshmi[2]
[1&2](Department of Geography, GCW Parade, Jammu, India)

Abstract
The agricultural sector plays a vital role in India's economy and supports over half of the population. Traditional methods of estimating crop production face challenges such as lack of timely information and data accessibility. Remote sensing, integrated with Geographic Information System (GIS), offers a solution by providing timely and accurate crop yield assessments. Landsat, a remote sensing sensor, offers suitable spatial resolution and spectral characteristics for monitoring agriculture. Remote sensing, coupled with GIS, can contribute to efficient crop statistics systems, including acreage estimation, yield estimation, and forecasting. Past studies have successfully used remote sensing for crop acreage estimation, employing various methodologies and data analysis techniques. Dhampur, a tehsil in Bijnor district, Uttar Pradesh, relies heavily on agriculture, particularly the cultivation of sugarcane, wheat, and rice. This paper highlights the significance of remote sensing in agricultural management and highlights the specific case of Dhampur as an example.
Keywords: *Crop acreage, Remote sensing, Principal Component Analysis (PCA), Dhampur*

Received 07 May, 2023; Revised 16May, 2023; Accepted 19 May, 2023

I. Introduction

Agriculture provides the principal means of livelihood for the major Indian population. Agriculture and allied sector contribute 20 per cent of GDP and provide support to more than 50% of population (Ministry of Finance, 2021).Early assessment of crop yield is of paramount importance for making decisions regarding agricultural problems and management. The traditional system of estimation of crop production has several problems i.e., lack of timely information, variation in statistical figures, accessibility and quick retrieval of data and heavy burden of work on village level workers, in comparison remote sensing not only solves these aforesaid problems but has several other advantages over the traditional system of crop estimation (Singh R and Goyal R.C 2000)

Remote sensing has a potential of estimating crop acreage at district, regional and national level. Remote sensing integrated with Geographical Information System (GIS) has becomes a very useful tool for the management of available agricultural resource. Recent developments such as sensor quality and processing advances has made remote sensing even stronger player for a country like India in efforts to estimate crop area and production, thus Remote sensing accompanied with GIS may lead to the development of an efficient integrated system for crop statistics like crop acreage estimation, crop yield estimation and crop yield forecasting.

Landsat sensor offers spatial resolution of 30 meter with a swath width of 185 km which provide appropriate spectral characteristics for agriculture monitoring. This sensor has a repeat coverage of 16 days which is very important to any vegetation related remote sensing analysis, and is particularly important for real time, within season crop acreage and or yield estimation. The 30-meter spatial resolution is sufficient for the accurate identification of large homogenous crop fields (Mueller R, 2004) Landsat offersseven spectral bands which acquire data in the visible green, visible red, near infrared (NIR), short wave infrared (SWIR) and thermal infrared bands.

Remote sensing has a long history of use in crop acreage estimation and assessment. Numerous scholars have used remote sensing for crop acreage estimation. (Michael E & Mitchell L. Graham1992) used landsat TM data for crop area estimation in· the Mississippi delta region. They used Battese-Fuller model to estimate crop acreage at country level. To assess classification and estimation accuracy methods such as regression determination coefficient, Relative efficiency and coefficient of variation were used. (Wu Bingfang & Li Qiangzi, 2003) provided a new methodology where two-stage sampling procedure was used to estimate crop acreage using Landsat TM data. (Mergerson J W, 1981) estimated crop area using ground-gathered and sampled landsat data in which he mainly focused on the sampling schemes of classified Landsat pixels. (Singh A et al., 2004) estimated potato acreage in the Indo-Gangetic plains using two-date WiFS data. NDVI model and Decision rule algorithm was used for potato crop discrimination.

Crop acreage estimation is generally accomplished through ground surveys supplemented with remotely sensed data. Dhampur is one among 5 tehsils in district Bijnor of state Uttar Pradesh, India. It is located at 29.30^{o} N Latitude and 78.50^{o} E Longitude. As of 2011 India census Dhampur had a population of 50997. Males constitute 53% of the population and females 47% respectively. Dhampur's economy is based on agriculture. Sugarcane, wheat and rice are mainly gown in all three cultivation seasons. Kharif season has the highest yield in comparison to rabi and zaid season owing to favorable climatic condition during this season.

Study Area

Dhampur is a sub division of district Bijnor of state Uttar Pradesh, India. It is located at 29.30^{o} N Latitude and 78.50^{o} E Longitude. The total area of the tehsil is 88,529 hectares. As of 2011 population census Dhampur had a population of 50,997. Males constitute 53% of the population and females 47% respectively. Dhampur's economy is based on agriculture. Sugarcane, wheat and rice are the main crops grown in both the cultivation seasons. KharifSeason has higher yield in comparison to Rabi season owing to favorable climatic condition. The average annual temperature ranges between 24^{0}C and 27^{0}C. The average annual rainfall of the tehsil remains less than 60 cms (Indian Meteorological Department, 2013). Net irrigated area of the tehsil accounts for 50,812 hectares. Different sources of irrigation like tube wells and canals are used for growing crops.

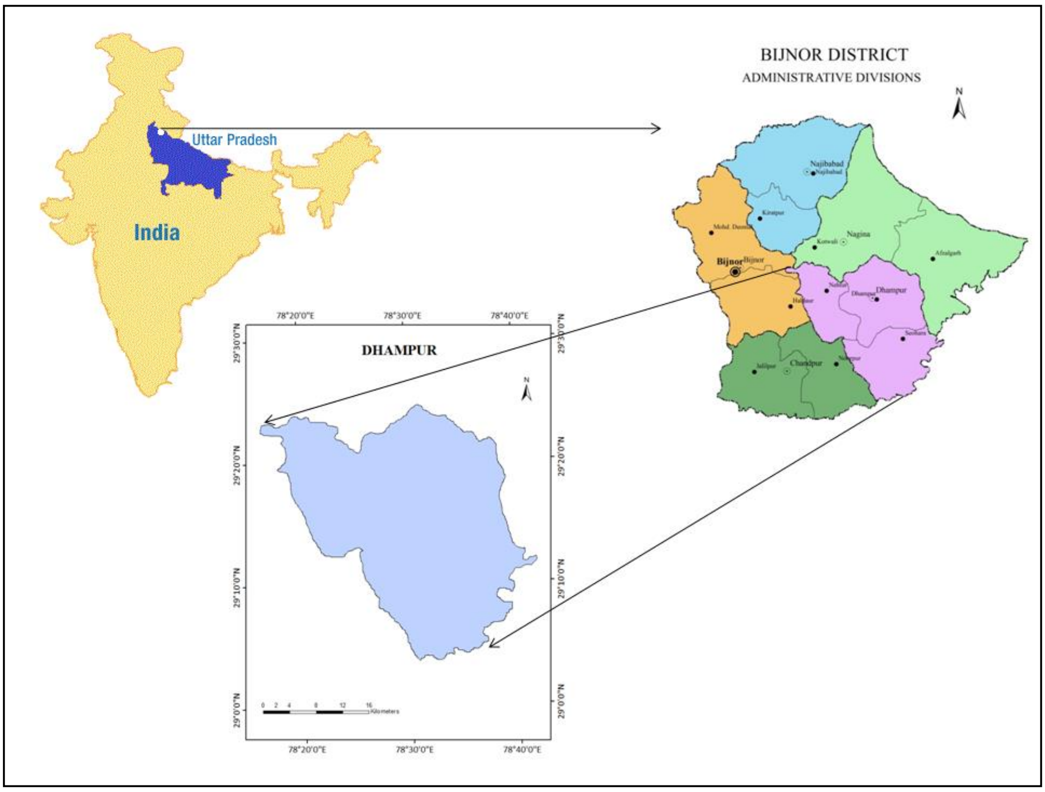


Figure 1. Location Map of Study area

II. Material and methods

Landsat 5 TM digital data has been used for identification and estimation of crop acreage of rabi and Kharif season. The administrative boundary of the study area was demarcated from Survey of India Topographical sheets on the scale of 1:250000 and overlaid on the image to extract the entire pixels of the study area. The statistical data of crop were collected from Land Record Office and economics and Statistics

Department, Government of Uttar Pradesh. Surveyed and GPS data were collected for ground truth requirements.

For acreage estimation of Dhampur Tehsil, we analyzed Principal component analysis (PCA) and supervised classification techniques. Principal Component Analysis (PCA) is a statistical technique used to reduce a set of correlated multivariate measurements to a smaller set where the features are uncorrelated to each other. The multispectral (or multi band) images have been acquired in different parts of the electromagnetic spectrum thus retaining correlation amongst bands. The techniques of PCA have been incorporated as a special transformation in digital image processing of satellite images where a number of correlated bands of the image data have been reduced to few uncorrelated bands. Whereas supervised classification is the method of land use and land cover classification based on ground truth. The results obtained through these models were compared with surveyed statistical data for evaluating the potential of each model.

Table i. Sensor characteristics

Sensor Characteristics	Landsat 5 (TM)
In Field of View (IFOV)	30 meters (Bands 1-5) 120 meters (Band 6)
Spectral Bands	Band 1: 0.45 - 0.52 (um) Band 2: 0.52 - 0.60 (um) Band 3: 0.63 - 0.69 (um) Band 4: 0.76 - 0.90 (um) Band 5: 1.55 - 1.75 (um) Band 6: 10.40 - 12.5 (um) Band 7: 2.07-2.35 (um)
Swath Width	185 km
Quantization	8 bits
Repeat Cycle	16 days

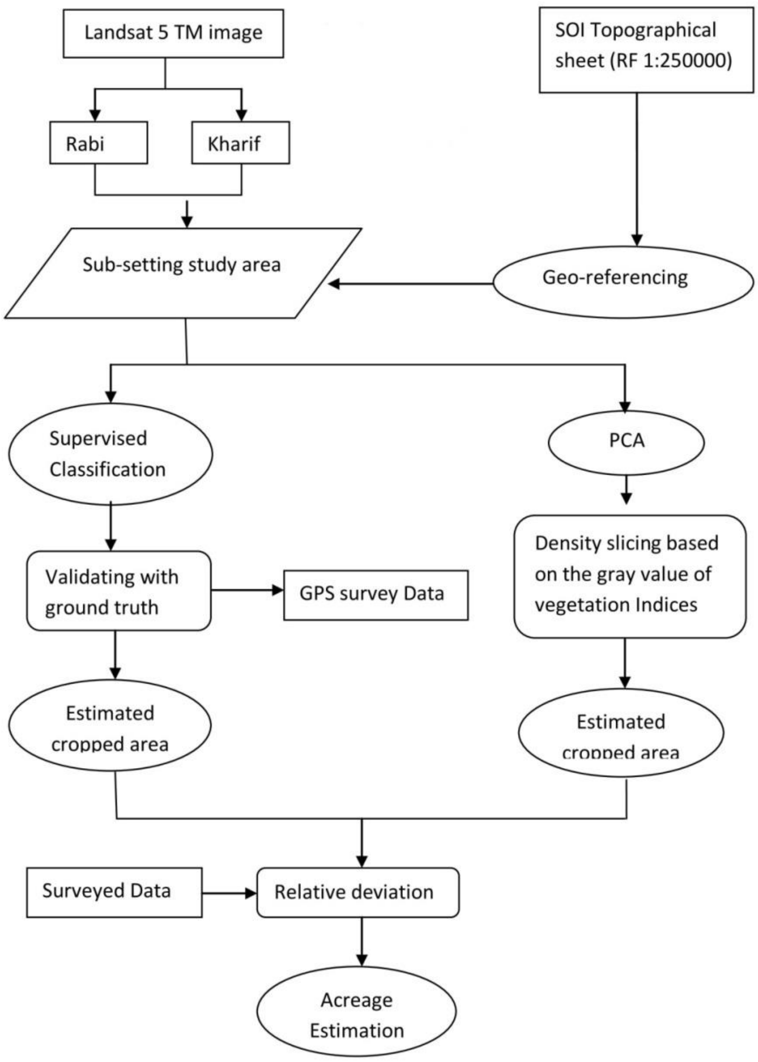
Error matrix was used to carry out the accuracy assessment of supervised classification. The matrix was calculated on category-by-category basis. The overall accuracy was computed by dividing the total number of correctly classified pixels by the total number of reference pixels. The producer accuracies were calculated by dividing the number of correctly classified pixels in each category by the number of training sets pixels of that category (the column total). User accuracies were computed by dividing the number of correctly classified pixels in each category by the total number of pixels classified in that category (the row total).

Further Kappa coefficient was used as a measure of agreement between model predictions and reality or to determine if the values contained in an error matrix represent a result significantly better than random (Jensen, 1996). The Kappa Coefficient is a discrete multivariate measure that differs from the usual measures of overall accuracy assessment in basically two ways. First, the calculation takes into account all of the elements of the error matrix, not just the diagonals of the matrix (Foody, 1992). This has the effect of taking into account chance agreement in the classification. The resulting Kappa measure compensates for chance agreement in the classification and provides a measure of how much better the classification performed in comparison to the probability of random assigning of pixels to their correct categories.

Kappa Coefficient (K_{hat}) Coefficient of Agreement was calculated by the following formula:

Kappa Coefficient (K_{hat}) $$K = \frac{N\sum_{i=1}^{k} xii - \sum_{i=1}^{k}(x_{i+} * x_{+i})}{N^2 - \sum_{i=1}^{k}(x_{i+} * x_{+i})}$$

Where N is the total number of sites in the matrix, r is the number of rows in the matrix, xii is the number in row i and column i, x+i is the total for row i, and xi+ is the total for column I. K_{hat}>0.80 represent strong agreement and good accuracy. 0.40-0.80 is moderate, <0.40 reflects poor agreement and low accuracy.



Relative deviation was computed to assess the accuracy of acreage estimation by PCA and supervised classification from the actual surveyed data. It also helped to assess the suitability of models for crop acreage estimation.

III. Results and discussion

PCA and supervised classification models were used to estimate crop acreage. These models were then compared to select the best model for acreage estimation. In PCA model the gray values of all pixels were classified into high, medium and low on the basis of digital number (DN) by density slicing method. On the other hand, supervised classification was performed based on the pixel brightness of image and GPS data. Gaussian Maximum likelihood classifier method was also applied for estimating crop acreage. On the basis of the reflectance samples pixels were selected from the image. Some pixels were wrongly classified due to the similarity of brightness value. These errors were minimized by the recoding of pixel of such locations on the basis of GPS data. Six distinct classes were identified as agriculture, other vegetation, settlement, fallow land, river and river bed.

Table ii. Area under different classes as calculated by PCA

Classes	Rabi		Kharif	
	Range of Gray Values	Area in hectares	Range of Gray Values	Area in hectares
Other Vegetation	0-155	2477.71	0-171	2005.36
Crops	156-167	23425.92	172-179	66357.57
Other LULC	168-255	65103.57	180-255	22196.38

Source: Authors' calculation based on PCA

Crop acreage was estimated through gray values of PCA image. The acreage estimated by PCA model for Rabi season was 26758.29 hectares whereas for Kharif season it was 69851.19 hectares. Crop acreage by supervised classification for rabi season was 23425.9 hectares, Whereas for Kharif season it was calculated as 66357.2 hectares

Table iii. Area under different land use and land cover.

Sno.	Class	Rabi Season	Kharif Season
1	Agriculture	23425.92	66357.57
2	Fallow land	45773.55	13103.12
3	Settlement	7208.64	5595.62
4	Other vegetation	2477.7	2005.36
5	River	2062.99	977.51
6	River bed	7580.69	514.37

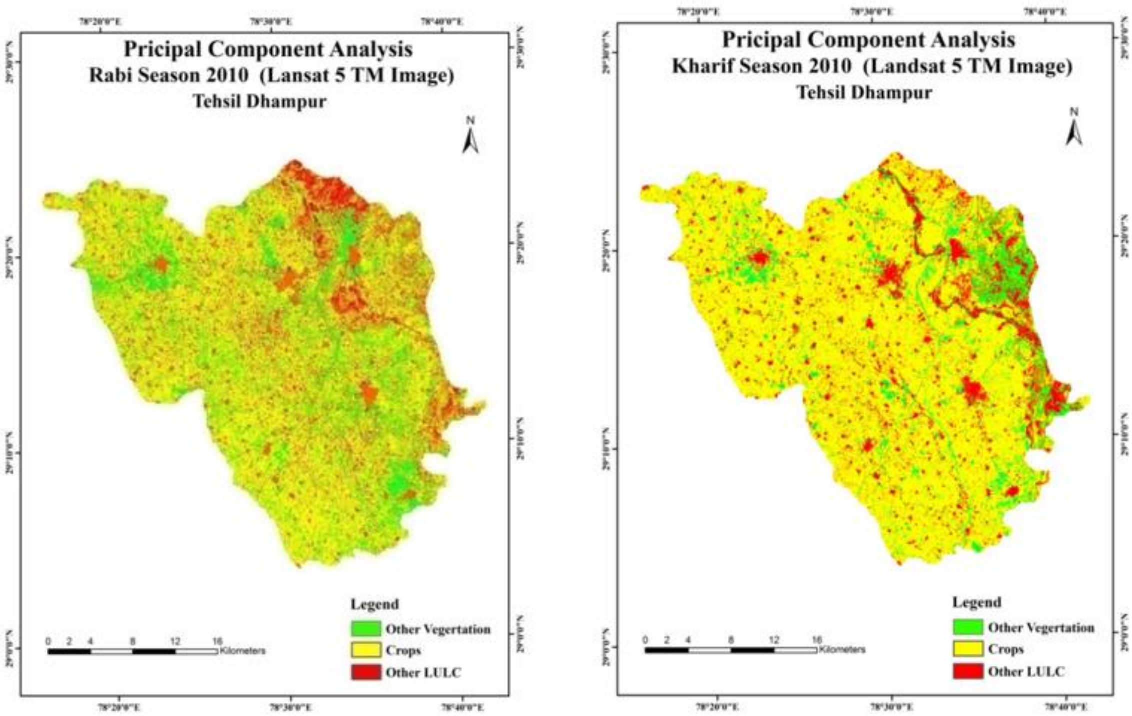


Fig iii. Principal Component Analysis of Rabi and Kharif season

Validation of the supervised classification

Accuracy assessments of supervised classification for both the seasons are calculated on the basis of producer, user and overall accuracy. The error occurred during sample pixel selection is examined through this accuracy assessment. The accuracy assessment Table 4 and 5 show the percentage of accuracy for classification of Rabi and Kharif season using Landsat 5 TM data. Overall accuracy of land use/ land cover classification of rabi season was 89.62%. Kappa coefficient of agreement showed an accuracy of 87.23%,while overall accuracy of Kharif season was calculated at 88.23% (Table 5). Kappa coefficient for Kharif season was 85.45%. Thus we can say that the pixels classified in all the six categories were very much closer to accuracy.

Table iv. Accuracy assessment of Rabi season.

Classes	Agriculture	Settlement	Fallow land	Other vegetation	River	River bed	Row Total	User's Accuracy
Agriculture	**90**	0	9	8	0	0	**107**	84.11%
Settlement	3	**93**	0	0	0	0	**96**	96.87%
Fallow land	0	0	**120**	5	0	7	**132**	90.90%
Other vegetation	8	0	0	**53**	0	0	**61**	86.88%
River	0	4	0	0	**41**	2	**47**	87.23%
River bed	0	0	0	0	6	**52**	**58**	89.65%
Column Total	**101**	**97**	**129**	**66**	**47**	**61**	**501**	
Producer's Accuracy	89.11%	95.87%	93.02%	80.30%	87.23%	85.24%		*89.62%

** Overall accuracy calculated from diagonal total i.e. 449*

Table v. Accuracy assessment of Kharif season.

Classes	Settlement	River bed	Other vegetation	River	Agriculture	Fallow land	Row Total	User's Accuracy
Settlement	**74**	0	0	6	0	5	**85**	87.05%
River bed	0	**48**	0	4	0	0	**52**	92.30%
Other vegetation	0	0	**53**	0	7	0	**60**	88.33%
River	4	5	0	**54**	0	0	**63**	85.71%
Agriculture	0	0	6	0	**130**	5	**141**	92.19%
Fallow land	5	0	0	0	9	**61**	**75**	81.33%
Column Total	**83**	**53**	**59**	**64**	**146**	**71**	**476**	
Producer's Accuracy	89.16%	90.56%	89.83%	84.37%	89.04%	85.91%		*88.23%

**Overall accuracy calculated from diagonal total i.e. 420*

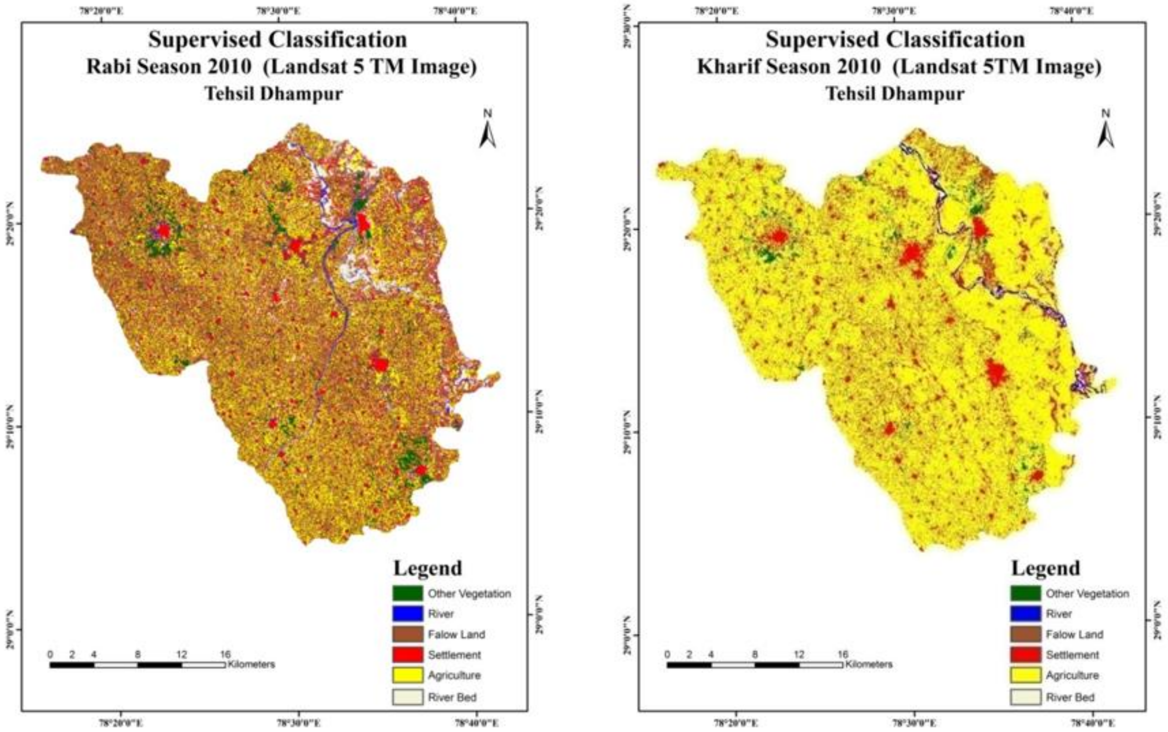


Fig iv. Supervised Classification of Rabi and Kharif season.

Validation of PCA and supervised classification based on surveyed data

The estimated crop area using PCA and supervised classification and their relative deviation is given in table 6. The table shows that in rabi season the relative deviations of PCA from the surveyed data is positively deviated i.e.10.42%. Whereas for Kharif season the relative deviation of PCA is 7.71%. Supervised classifications for rabi season is negatively deviated from surveyed data i.e. -2.31% on the other hand supervised classification for Kharif season is positively deviated i.e. 2.85% from the surveyed data.

Table vi. Relative deviation (RD) of estimated crop acreage from surveyed data (Area is in hectares)

Season	Surveyed Area	Supervised Classification	RD%	PCA	RD%
Rabi	23968	23425.9	-2.31	26758.29	10.42
Kharif	64460	66357.2	2.85	69851.19	7.71

IV. Conclusion

Single date Landsat 5 TM digital data was analyzed to estimate the crop acreage of Dhampur Tehsil. PCA and maximum likelihood supervised classification supplemented with ground truth were used to estimate the crop acreage. The results obtained through these models were compared with surveyed statistical data for evaluating the potential of each model. Accuracy assessment was carried by error matrix, further Kappa coefficient was also used as a measure of agreement between model prediction and reality.Analysis of models showed variation in results. Supervised classification using landsat 5 TM data showed lowest deviation in the estimated area from the surveyed data. Thus we conclude that supervised classification model with validation of GPS data can effectively provide accurate estimation of crop acreage in comparison to PCA model for better agricultural land use planning.

References

[1]. Foody, Giles M, 1992. On the Compensation for Chance Agreement in Image Classification Accuracy Assessment," Photogrammetric Engineering and Remote Sensing, vol 58, No. 10, 1459-1460 pp.

[2]. Jensen, J.R., 1996. Introductory Digital Image Processing: A Remote Sensing Perspective, Prentice Hall Series in Geographic Information Science, Prentice Hall, Upper Saddle River, New Jersey, 316 pp.

[3]. Mergerson J W, 1981. Crop Area Estimation Using Ground-Gathered and Sampled Landsat Data. United States Department of Agriculture, Economics and Statistics Service, Statistical Research Division ESS staff report.

[4]. Michael E & Mitchell L. Graham, 1992. Improved Crop Area Estimation in· the Mississippi Delta Region Using Landsat TM Data, National Agricultural Statistics Service, Presented at the ASPRS/ACSM/RT 92 Convention, Washington, D.C., August 3-7,

[5]. Ministry of Finance, 2021. Mid-Year Economic Analysis, Department of Economic Affairs, Economic Division, Government of India.

[6]. Mueller R, 2004. Assessment of TM & AWiFS imagery for cropland classification: three case studies, United States Department of Agriculture, National Agricultural Statistics Service Research and Development Division (NASS)

[7]. Singh A, 2004. Prospect of Potato Acreage Estimation Using Temporal WiFS Data in Indo-Gangetic Plains, The Indian Journal of agricultural sciences vol. 1 No. 2 7-12 pp

[8]. Wu Bingfang & Li Qiangzi. (2003). Crop area estimation using remote sensing on two stage stratified sampling. Institute of Remote Sensing Applications, Chinese Academy of Sciences, Beijing, China. Commission VI, WG VII/3.

[9]. Web source: http://indiainbusiness.nic.in/newdesign/index.php?param=economy_landing/213/2 [28th march 2015]