Int.J.Curr.Microbiol.App.Sci (2022) *11*(05): 115-124

EXCELLENT PUBLISHERS

International Journal of Current Microbiology and Applied Sciences
ISSN: 2319-7706 **Volume 11 Number 05 (2022)**
Journal homepage: http://www.ijcmas.com

Original Research Article https://doi.org/10.20546/ijcmas.2022.1105.015

Evaluation of Machine Learning Classifiers for Crop Classification – A Case Study of Veppanthattai Taluk, Perambalur District

R. Sakthivel[1] and B. Linda Theres[2*]

[1]*Department of Geography, Guindy Campus, University of Madras, Chennai – 600025, India,*
[2]*Department of Civil Engineering, SASTRA Deemed to be University, Thanjavur – 613401, India*

**Corresponding author*

Keywords

Crop classification, Maximum Likelihood Classifier, Minimum Distance to Mean Classifier

Article Info

Received:
10 April 2022
Accepted:
29 April 2022
Available Online:
10 May 2022

A B S T R A C T

India is a country that has major dependence on agriculture. In spite of all the careful measures, there is a loss of agricultural land due to improper use of land. One way to tackle the issue is hand, is by monitoring and quantifying the crop. Also, Crop discrimination and Acreage estimation are essential for planning and policymaking at various administrative levels. In earlier days, a field-based survey was executed with meddlesome factors like time consumption, and labor-intensive. But with remote sensing and GIS technology in the picture, the survey became a cakewalk. It can provide timely and accurate crop inventory information. Also, with a focus on Machine Learning in recent years, crop classification became more precise. The main objective of this study is to evaluate the performance of Maximum Likelihood Classification (MLC), Minimum Distance to Mean classification (MDMC), Support Vector Machine (SVM) classification, and Neural Net (NN) classification for crop discrimination and acreage estimation in Veppanthattai Block of Perambalur District in Tamil Nadu using Sentinel 2B. Ground truth data were collected randomly over the study area, in that 60% of data used for training and 40% data was used for testing the classification. It was inferred that MLC was the best with a kappa coefficient of 0.89, compared with MDMC, SVM, and NN which exhibited poor classification results with a kappa coefficient of 0.47, 0.85, and 0.83 respectively.

Introduction

Agriculture stands as the backbone of the Indian economy as almost two-thirds of the population is involved. It imparts major influence as it also delivers raw materials for many manufacturing units. Hence hawk-eyed management of the sector is a necessity. Crop production estimation involves the determination of the total area under crop and prediction of the yield per unit area. To aid the purpose, Remote sensing techniques have demonstrated their potentiality in providing information on the characteristics and spatial distribution of agricultural resources because of their unique advantages of providing multi-spectral, multi-temporal and multi-spatial resolutions (Zhang *et al.*, 2011). Use of satellite remote sensing data has also proved to be more cost-effective, reliable, timely, and faster than the conventional ground-based surveys of agricultural resources (Navalgund *et al.,* 1991; Patel *et al.,* 2008; Panigrahy *et al.,* 2009). A crop information system is one way that remote sensing can provide valuable information to decision-makers. Create a database of the crops, is required to perform crop classification and acreage estimation (Dhumal *et al.,* 2013; Howard *et al.,* 2012; Chen *et al.,* 2012; Tu and Shang, 2018; Wang *et al.,* 2019). Satellite images coupled with several machine learning-based classifiers were used for the application (Tateishi *et al.,* 1991; Friedl and Brodley, 1997; Hastings, 1997; Friedl *et al.,* 1999). To be able to increase the accuracies of crop identification and area estimation, we need to have a better understanding of the crop and the underlying soil characteristics that influence the radar backscatter throughout the growing season and identify a suitable methodology to extract crop information from optical satellite imagery (Dadhwal *et al.,* 1991; Panigrahy *et al.,* l99l; Mosleh *et al.,* 2015; Maurya, 2011; Aboelghar, 20l3; Mulianga *et al.,* 2013; Prasad, *et al.,* 2006). With the above-mentioned objectives in mind, the study was executed to evaluate classifiers for crop discrimination and acreage estimation. Though there is no one universally acceptable classifier, several kinds of research were performed to identify the best classifier that fit the purpose (Linda and Selvakumar, 2022). A study was undertaken for the same area using SAR data (Sakthivel *et al.,* 2021). But this paper discusses performance classifiers with multispectral data.

Study Area

Veppanthattai block of Perambalur District is a centrally located inland area of Tamil Nadu. This area is located between 11°16' N to 11°31' N and 78°38' E to 79°0' E covering an area of 580 sq.km (Figure1). It has an agricultural area of 432 sq.km. This Veppanthattai block has 39 revenue villages. The mean annual rainfall of the area is 908 mm, which that 475 mm received by North East Monsoon and 314 mm received by South West Monsoon. The climate is hot sub-humid to semi-arid. The mean annual maximum and minimum temperature are 32.60˚C and 22.20˚C respectively. Black cotton soil, clay loam, and red sandy soil are the predominant soil types. In this area, major crops are grown such as Cotton, Maize, Paddy, Tapioca, Sorghum, Onion, Groundnut, and Sesamum in which cotton and maize are the dominant crops**.** In this study, four major crops namely Cotton Maize Paddy, and Tapioca were emphasized for crop classification.

Materials and Methods

The process flow to achieve the goals of the study is given in figure2.The sentinel 2B from the peak vegetation period was used for crop classification since it provides promisingly better resolution than the Landsat dataset (Orynbaikyzy *et al.,* 2022; Tuvdendori *et al.,* 2022; Khuong *et al.,* 2022; Seydi *et al.,* 2022). The machine learning classifiers namely Maximum Likelihood Classification, Minimum Distance to Mean Classification, Support Vector Machine Classification, and Neural Net Classification were analyzed for its performance in discriminating the crops under study. Finally, the classifier is identified that is the best fit for the application (Meshram *et al.,* 2022; Ghayour *et al.,* 2021; Aicha and Abderrahman, 2021; Erdaney *et al.,* 2022; Parida *et al.,* 2021).

Ground Truth Data Collection

To train and test the classifiers, ground truth data were collected throughout the Veppanthattai block, when it was at its peak vegetative stage. For this purpose, a mobile application namely Field Area Measure was used. Among the data collected, 60% was used for training and 40% for testing the classifiers to measure their performance.

Pre-processing

Any multispectral data will suffer errors due to atmospheric interaction. Hence the procured data was done with atmospheric correction using SNAP software. The corrected data were then resampled to the same cell size to avoid conflicts in further processing. Then the layers were stacked and clipped to the extract region of interest under study. Finally, by applying masks through manual digitization, agricultural land within the study area extent was extracted.

Maximum Likelihood Classification

Maximum likelihood classification assumes that the statistics for each class in each band are normally distributed and calculates the probability that a given pixel belongs to a specific class. Unless you select a probability threshold, all pixels are classified. Each pixel is assigned to the class that has the highest probability (that is, the maximum likelihood). If the highest probability is smaller than the threshold you specify, the pixel remains unclassified.

Minimum Distance to Mean Classification

The minimum distance technique uses the mean vectors of each end member and calculates the Euclidean distance from each unknown pixel to the mean vector for each class. All pixels are classified to the nearest class unless a standard deviation or distance threshold is specified, in which case some pixels may be unclassified if they do not meet the selected criteria. Minimum distance classifiers are direct in concept and implementation but are not widely used in remote sensing work. In its simplest form, minimum distance Classification is not always accurate; there is no provision for accommodating differences in variability of classes, and some classes may overlap at their edges.

Support Vector Machine Classification

Support Vector Machine (SVM) is a supervised classification method derived from statistical learning theory that often yields good classification results from complex and noisy data. It separates the classes with a decision surface that maximizes the margin between the classes. The surface is often called the optimal hyperplane, and the data points closest to the hyperplane are called support vectors. The support vectors are the critical elements of the training set. 21 You can adapt SVM to become a nonlinear classifier through the use of nonlinear kernels. While SVM is a binary classifier in its simplest form, it can function as a multiclass classifier by combining several binary SVM classifiers (creating a binary classifier for each possible pair of classes).

Neural Net Classification

The feed forward neural network technique uses the back propagation algorithm for supervised learning. The number of hidden layers is user-defined. There is a provision in choosing between a logistic or hyperbolic activation function. Learning occurs by adjusting the weights in the node to minimize the difference between the output node activation and the output. The error is back propagated through the network and weight adjustment is made using a recursive method. You can use Neural Net classification to perform non-linear classification.

Accuracy Assessment

The performance of all the classifiers was tested using ground truth data. The error matrix denoting the agreement and disagreement (Lillesand *et al.,* 2007) was developed and numerically evaluated by calculating overall accuracy (Congalton, 1991). The overall accuracy determines the percentage of correctly classified pixels as a whole. And estimated using the following equations (Equation 1)

$$Overall\ accuracy = \frac{\sum(correctly\ classified\ classes\ along\ the\ diagonal)}{\sum(Row\ total\ or\ Column\ Total)} \quad \text{...Eqn 1}$$

Also, the Kappa coefficient which is a measure of proportional improvement by the classifier over a purely random assignment to classes was estimated using the formula as given in equation 2.

$$k = \frac{NA - B}{N^2 - B} \quad \text{...Eqn 2}$$

Where, k denotes kappa coefficient, N denotes a total number of testing sites, A is the sum of elements along the diagonal and B denotes the product of row total and column total. All the above-mentioned performance measures were computed for all the classifiers in the discussion.

Results and Discussion

Classification

After the pre-processing of Sentinel 2B data in SNAP software. The satellite data was clipped to the Veppanthattai block. As the initial step mask is applied to extract the agricultural extent. This extracted region of interest is then used to classify into Cotton, Maize, Paddy, Tapioca, and other crops. After carefully studying the crops, the ground truth data were collected for the major crops under study.

These were used for training and testing the classifiers. The classification was performed using MLC, MDMC, NN, and SVM algorithms, and the output was mapped as shown in Figures3, 4, 5, and 6. The classified data is then converted to vector format to compute the areal extent of each crop in hectares (Table1).

Evaluation of Classifiers

The classified results are then tested using ground truth data. The performance measures of the classifiers used for the study undertaken were overall accuracy and kappa coefficient. They are computers for all the four classifiers to be analyzed and the consolidated results are shown in table 2. MLC has the best overall accuracy of 92.69 with a kappa coefficient of 0.89. Comparatively, MDMC shows the poor result with the lowest kappa coefficient of 0.47, as the algorithm is simple and failed to comply with the complexity of the spectral reflectance data of each crop. In conclusion, though NN and SVM are fairly acceptable with their performance, it was inferred that MLC, is the fittest classifier for the region of interest and application under study.

Fig.1 Location map of Veppanthattai block in Perambular district

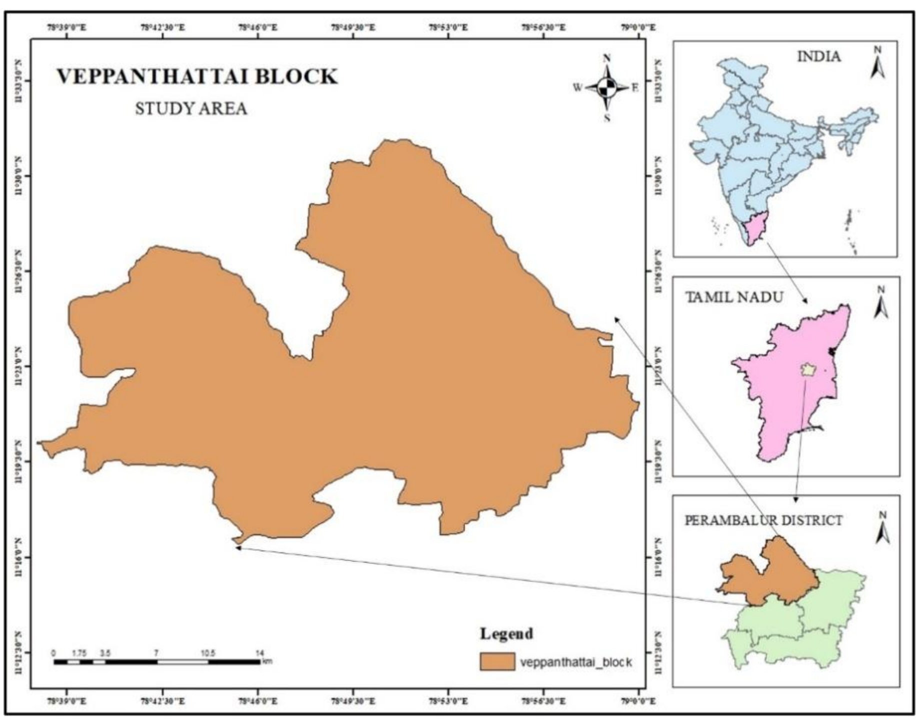


Table.1 Acreage estimation of each crop in hectares

Classification Algorithm	Acreage Estimation (Hectares)				
	Cotton	Maize	Paddy	Tapioca	Other crops
Maximum Likelihood Classification	7876	18990	8314	120	7900
Minimum Distance Classification	15556	8465	11436	2707	5036
Neural Net Classification	17317	17367	5369	148	2999
Support Vector Machine Classification	12854	15667	6665	416	7598

Fig.2 Methodology flowchart

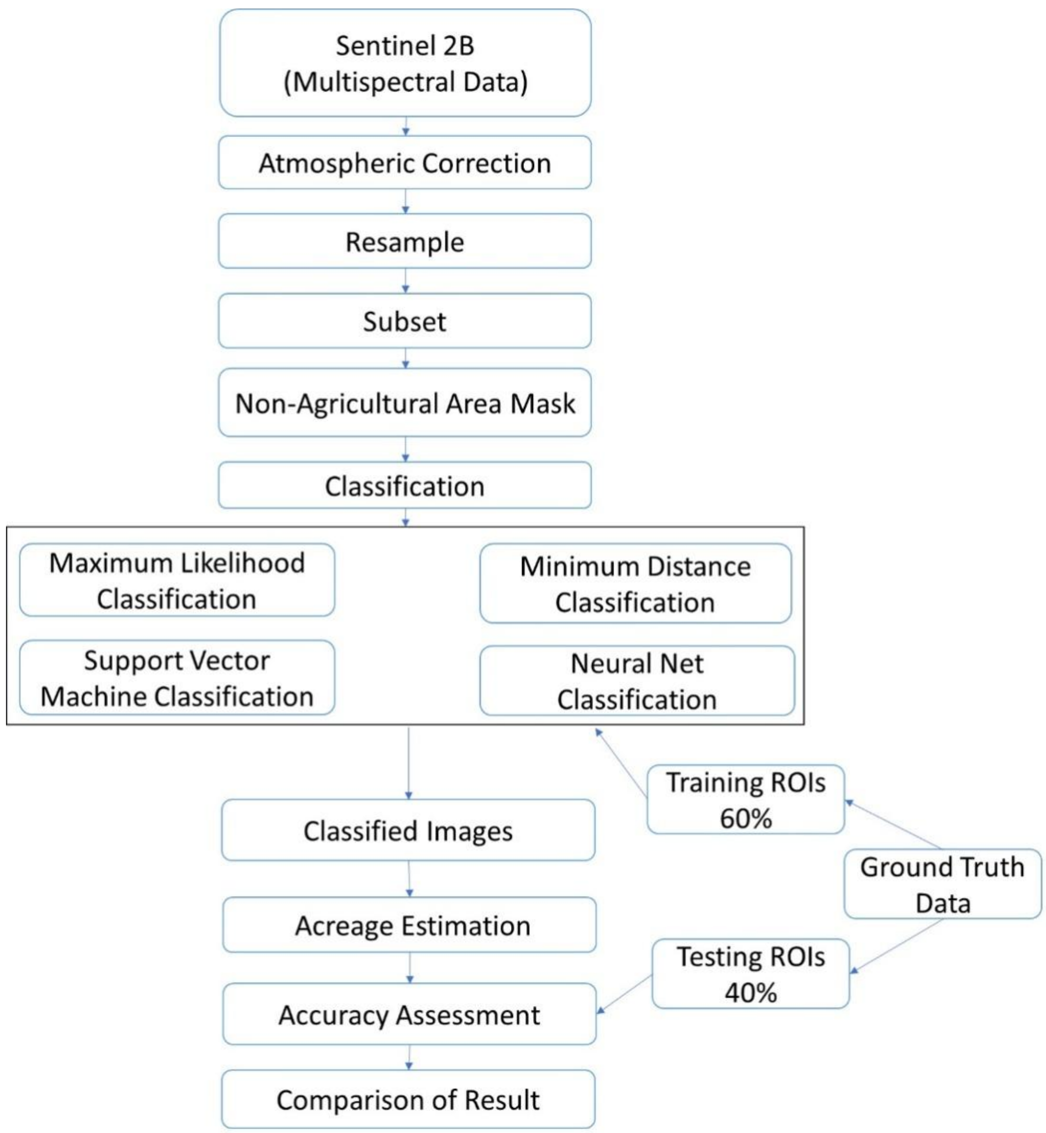


Table.2 Performance measures of the classifiers

Classification Algorithm	Overall Accuracy	Kappa coefficient
Maximum Likelihood Classification	92.69	0.89
Minimum Distance Classification	60.03	0.47
Neural Net Classification	88.52	0.83
Support Vector Machine Classification	89.96	0.85

Fig.3 The Classified Image using Maximum Likelihood Classification

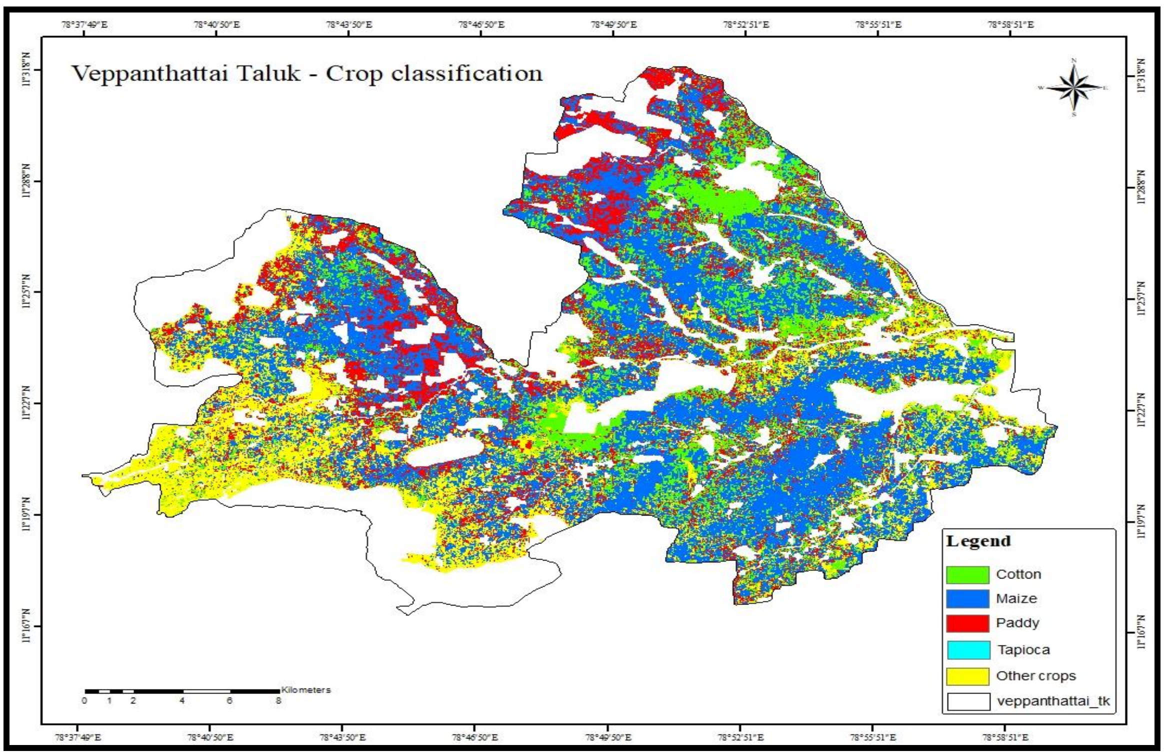


Fig.4 The Classified Image using Minimum Distance Classification

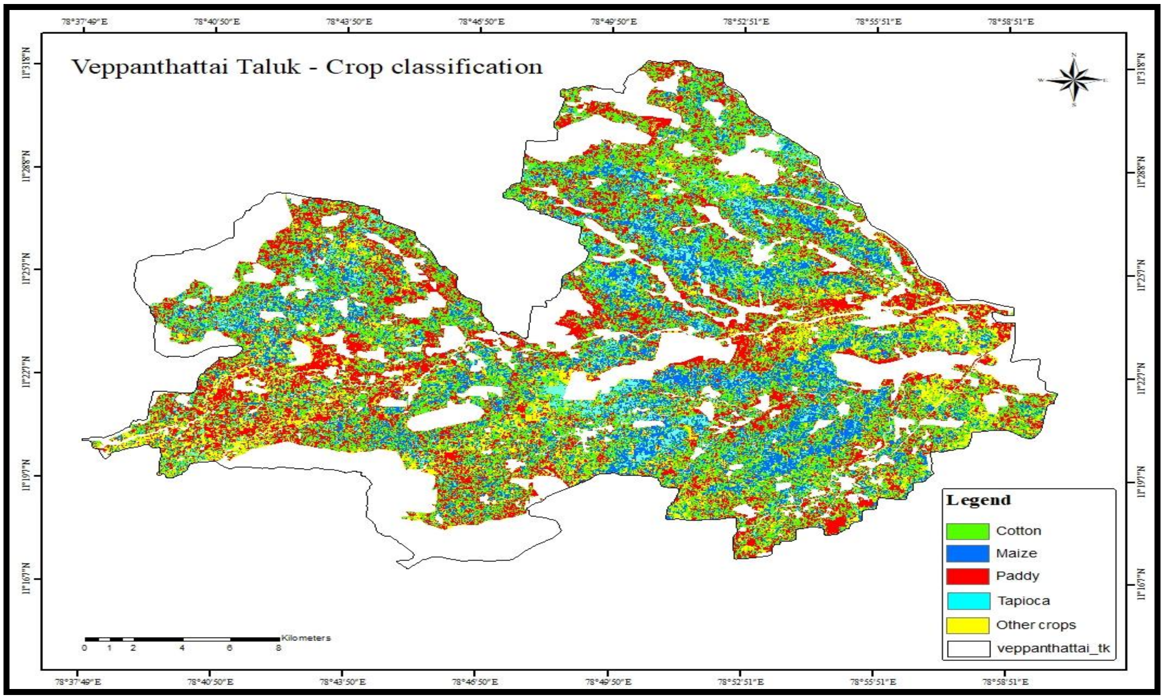


Fig.5 The Classified Image using Neural Net Classification

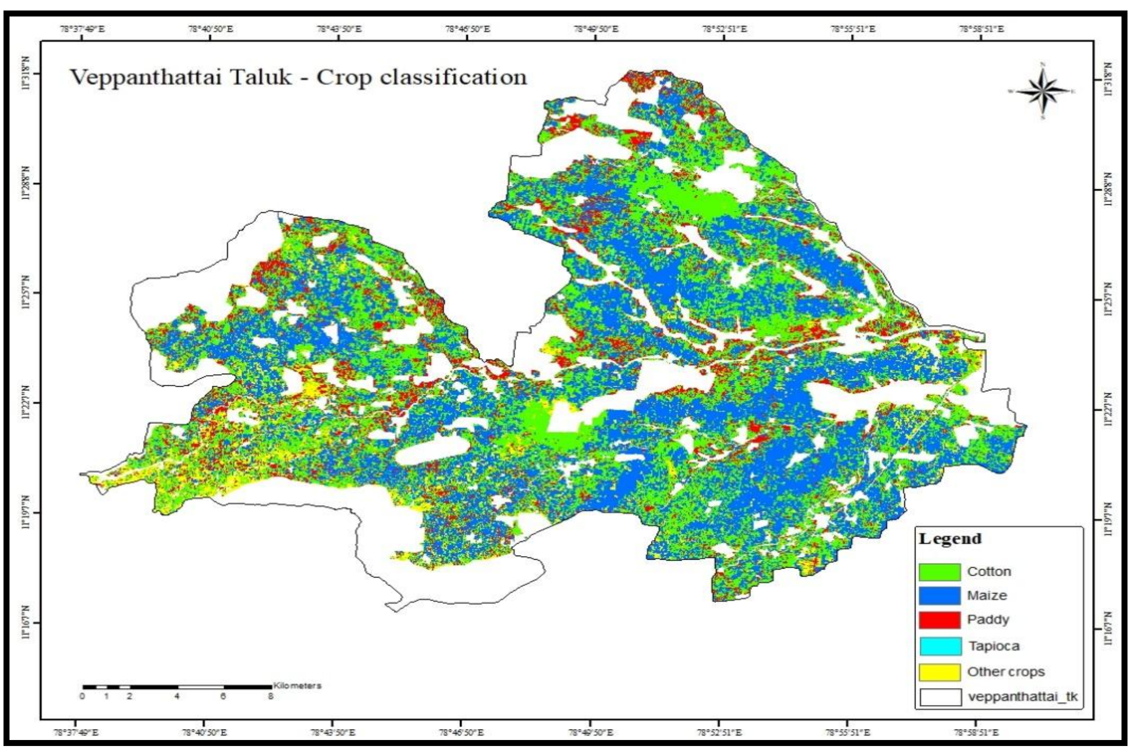
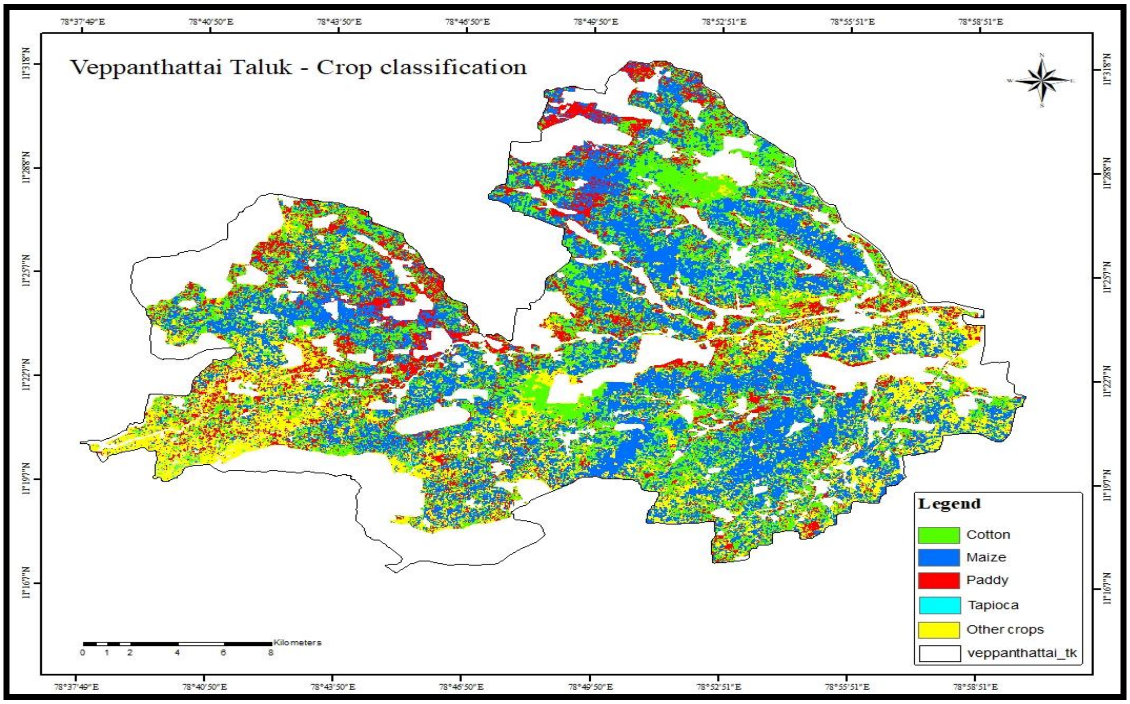


Fig.6 The Classified Image using Support Vector Machine Classification



In crop management, it is crucial to have means to aid in acreage estimation. The geographical information system with its ability to handle spatial entities and its geometry, along with computing technology with machine learning algorithms, the application under study has become an easy task. It is known that In the Classification of crops optical data perform better than the SAR data because it has a higher spectral and spatial resolution. This study aimed to compare four classifiers with Sentinel 2B data and inferred that the Maximum Likelihood Classification algorithm performs better than other classifiers such as Support Vector Machine, Neural Net, and Minimum Distance Classification.

References

Zhang, H., Lan, Y., Lacey, R., Hoffmann, W. C., Westbrook, J. K., 2011. Spatial analysis of NDVI readings with different sampling densities. Transactions of the ASABE 54(1), 349-354.

Tateishi, R., K. Kajiwara, and T. Odajima, 1991. Global land cover classification by phenological methods using NOAA GVI data. AsianPacific Remote Sensing Journal. vol. 4, no. 1, pp. 41-50.

Friedl, M. A., and Brodley, C. E., 1997. Decision tree classification of land cover from remotely sensed data. Remote Sensing of Environment. 61, 399-409.

Hastings, David A., 1997. Land cover classification: some new techniques, new source data. Proceedings, 18th Asian Conference on Remote Sensing. Tokyo, Asian Association of Remote Sensing. Pp. JS-2-1 to JS-2-6.

Friedl, M. A., Brodley, C. E., and Strahler, A. H., 1999. Maximizing land cover classification accuracies produced by decision tree at continental to global scales. IEEE Transactions on Geoscience and Remote Sensing. 37, 969-977.

Howard, D. M., Wylie, B. K., Tieszen, L. L., 2012, Crop classification modelling using remote sensing and environmental data in the Greater Platte River Basin, USA, International Journal of Remote Sensing, vol. 33, no. 19, pp. 6094-6108.

Chen, S., Zhao, Y, Shen, S., Crop classification by remote sensing based on spectral analysis, March 2012, Nongye Gongcheng Xuebao/Transactions of the Chinese Society of Agricultural Engineering, vol. 28, no. 5, pp. 154-160.

Yu, B.; Shang, S. Multi-Year Mapping of Major Crop Yields in an Irrigation District from High Spatial and Temporal Resolution Vegetation Index. Sensors, 2018, vol.18, pp. 3787.

Wang, L.; Dong, Q.; Yang, L.; Gao, J.; Liu, J. Crop Classification Based on a Novel Feature Filtering and Enhancement Method. Remote Sens., 2019, vol. 11, pp. 455.

Dhumal, R. K., Rajendra, Y. D., Kale, K. V., Mehrotra, S. C., Classification of Crops from remotely sensed Images: An Overview, International Journal of Engineering Research and Applications (IJERA) ISSN: 2248-9622 vol. 3, no. 3, May-Jun 2013, pp. 758-761.

Patel, N. R., Bhattacharjee, Mohammed, A. J., Tanupriya, B., Saha, S. K., 2006b. Remote sensing of regional yield assessment of wheat in Haryana, India. International Journal of Remote Sensing, 27, 19, pp. 4071-4090.

Panigrahy, R. K., Ray, S. S. & Panigrahy, S., 2009. Study on utility of IRS-P6 SWIR band for crop discrimination and classification. Journal of Indian Society of Remote Sensing. 37, 2, pp. 325-333.

Navalgund, R. R., Parihar, J. S., Ajai and Rao, P. P. N., 1991 Crop inventory using Remotely Sensed Data. Current Science 6l, pp. 162-171.

Dadhwal, V. K., Rahul, D. S., Medhavy, T. T., Jarwal, S. D., Khera, A. P., Singh, J., Sharma, T. and Parihar, J. S., 1991 Wheat Acreage Estimation for Haryana Using Satellite Digital Data. Journal of the Indian Society of Remote Sensing 19, pp. 1-15.

Panigrahy, S., Parihar, J. S., Patel, N. K., Dafhwal, V. K., Medhavy, T. T., Ghose, B. K., Ravi, N., Pani, K. C., Panigrahy, B. K., Sridhar, V. N., Mohanty, R. R., Nanda, S. K., Tripathy, D. P., Mishra, P. K., Bhatt, H. P., Oza, S. R., Sudhakar, S., Sudha, K. S., Kumar, P. and Das, N. K. 1991 Rice-Acreage Estimation for Orrisa Using Remotely Sensed Data. - Journal of the Indian Society of Remote Sensing 19, pp. 17 -26.

Mosleh, M. K., Hassan, Q. K. and Chowdhury E. H., 2015 Application of Remote Sensors in Mapping Rice Area and Forecasting Its Production: A Review. Sensors 15, pp. 769-791.

Maurya, A. K., 2011 Estimation of Acreage & Crop Production through Remote Sensing & GIS Technique. Geospatial world forum. pp l8-21.

Aboelghar, M., Arafat, S., Farag, E., 2013 Hyper Spectral Measurements as a Method for Potato Crop Characterization. International Journal of Advanced Remote Sensing and GIS 2,pp. 122-129.

Mulianga, B., Begue, A., Simoes, M. and Todoroff, P., 2013 Forecasting Regional Sugarcane Yield Based on Time Integral and Spatial Aggregation of MODIS NDVI. Remote Sensing 5,pp.2184-2199.

Prasad, A. K., Chai, L., Singh, R. P. and Kafatos, M., 2006 Crop yield estimation model for lowa using remote sensing and surface parameters. International Journal of Applied Earth Observation and Geoinformation 8, pp. 26-33.

Orynbaikyzy A, Gessner U, Conrad C, 2022. Spatial Transferability of Random Forest Models for Crop Type Classification Using Sentinel-1 and Sentinel-2. Remote Sensing 14(6):1493. https://doi.org/10.3390/rs14061493

Tuvdendorj B, Zeng H, Wu B, Elnashar A, Zhang M, Tian F, Nabil M, Nanzad L, Bulkhbai A, Natsagdorj N, 2022. Performance and the Optimal Integration of Sentinel-1/2 Time-Series Features for Crop Classification in Northern Mongolia. Remote Sensing. 2022; 14(8):1830. https://doi.org/10.3390/rs14081830

Khuong H. Tran, Hankui K. Zhang, John T. McMaine, Xiaoyang Zhang, Dong Luo, 2022. 10 m crop type mapping using

Sentinel-2 reflectance and 30 m cropland data layer product, International Journal of Applied Earth Observation and Geoinformation, Volume 107, 102692, https://doi.org/10.1016/j.jag.2022.102692.

Seydi S T, Amani M, Ghorbanian A, 2022. A Dual Attention Convolutional Neural Network for Crop Classification Using Time-Series Sentinel-2 Imagery. Remote Sensing. 2022; 14(3):498. https://doi.org/10.3390/rs14030498

Meshram, P. K., Rawat, K. S., Kumar, S., Singh, S. K., 2022. Crop-Type Classification Using Sentinel-2A and in Situ Data: Case Study of Shri Dungargarh Taluk of Rajasthan, India. In: Kolhe, M.L., Jaju, S.B., Diagavane, P.M. (eds) Smart Technologies for Energy, Environment and Sustainable Development, Vol 2. Springer Proceedings in Energy. Springer, Singapore. https://doi.org/10.1007/978-981-16-6879-1_18

Ghayour L, Neshat A, Paryani S, Shahabi H, Shirzadi A, Chen W, Al-Ansari N, Geertsema M, Pourmehdi Amiri M, Gholamnia M, Dou J, Ahmad A., 2021. Performance Evaluation of Sentinel-2 and Landsat 8 OLI Data for Land Cover/Use Classification Using a Comparison between Machine Learning Algorithms. Remote Sensing. 2021; 13(7):1349.

Aicha Moumni, Abderrahman Lahrouni, 2021. Machine Learning-Based Classification for crop-type mapping using the fusion of High – Resolution Satellite Imagery in Semiarid area. Scientifica, vol. 2021, Article ID 8810279, 20 pages.

Parida, B. R., Kumar, A. & Ranjan, A. K, 2021. Crop Types Discrimination and Yield Prediction Using Sentinel-2 Data and AquaCrop Model in Hazaribagh District, Jharkhand. KN Journal of Cartography and Geographic Information. (2021). https://doi.org/10.1007/s42489-021-00073-4

Erdanaey E, Kappas M, Wyss D, 2022. The identification of Irrigated Crop Types using Support Vector Machine, Random Forest and Maximum Likelihood Classification Methods with Sentinel – 2 Data in 2018: Tashkent Province, Uzbekistan. International Journal of Geoinformatics, Vol. 18, No. 2, ppL 37-53.

Dash, J., A. Mathur, G. M. Foody, P. J. Curran, J. W. Chipman & T. M. Lillesand, 2007. Land cover classification using multi-temporal MERIS vegetation indices, International Journal of Remote Sensing, 28:6, 1137-1159

Congalton, R. G., 1991. A Review of Assessing the Accuracy of Classifications of Remotely Sensed Data. Remote Sensing of Environment, 37, 35-46.

Linda Theres B, Selvakumar R, 2022. Comparison of landuse/landcover classifier for monitoring urban dynamics using spatially enhanced Landsat dataset. Environmental Earth Sciences 81, 142 (2022). https://doi.org/10.1007/s12665-022-10242-x

Sakthivel R, Venkatesh K, Parthipan S, 2021. Crop Discrimination and Acreage estimation of major crops for Veppanthattai Taluk, Perambalur District using Multi Temporal Sentinel 1A SAR data. International Journal of Engineering Research and Applications Vol. 11, Issue 10, pp: 14-20.

How to cite this article:

Sakthivel, R. and Linda Theres, B. 2022. Evaluation of Machine Learning Classifiers for Crop Classification – A Case Study of Veppanthattai Taluk, Perambalur District. *Int.J.Curr.Microbiol.App.Sci.* 11(05): 115-124. doi: https://doi.org/10.20546/ijcmas.2022.1105.015