

IMPROVED IN-SEASON CROP CLASSIFICATION PERFORMANCE USING ENSEMBLE LEARNING TECHNIQUE: A CASE STUDY OF LEKODA INSURANCE UNIT, UJJAIN, MADHYA PRADESH

Varun Pandey*¹, Karun Kumar Choudhary¹, C.S. Murthy¹, M.K. Poddar²

¹Agriculture Sciences and Application, National Remote Sensing Center, Hyderabad

²Agriculture Insurance Company of India Limited, New Delhi

Commission III, WG III/10

KEY WORDS: Crop Classification, Ensemble Learning, Random Forest, Naïve Bayes, Kappa coefficient.

ABSTRACT:

The classification of agricultural crop types is an important application of remote sensing. With the improvement in spatial, temporal and spectral resolution of satellite data, a complete seasonal crop growth profile and separability between different crop classes can be studied by using ensemble-learning techniques. This study compares the performance of Random Forest (RF), which is a decision tree based ensemble learning method and Naïve Bayes (a probabilistic learning technique) for crop classification of *Lekoda* gram panchayat, *Ujjain* district, using multi-temporal Sentinel 2 of Rabi 2017-18. The study area contains seven different classes of crop types, and in each class, we have used 65% of the ground data for training and 35% to test the classifier. The performance of RF classifier was found to be better than NB classifier. Kappa coefficient of RF classifier in mid of the crop season (December-January) was found to be 0.93. This result indicates that an accurate in-season crop map of the study area can be generated through integrated use of Sentinel 2 temporal data and RF classifier.

1. INTRODUCTION

Accurate spatial crop map is an essential input for a number of agricultural applications like crop insurance coverage assessment, crop yield estimation, crop monitoring & planning etc. Multispectral satellite data is being used worldwide to map specific crop and its area assessment (Bahram Salehi *et. al*). With the improvement in spatial, temporal and spectral resolution of satellite data for *e.g.* Sentinel 2, which is available at a temporal resolution of 5 days with 13 bands at different spatial resolution of 10m, 20m and 60m, a complete crop growth profile and separability between different crop classes can be studied. The challenge lie in the analysis of voluminous data associated with it. Such voluminous data can be analysed by using techniques of ensemble learning. Ensemble learning uses the combination of multiple models, such as classifiers, to produce a single predictive model. The resulting model, referred to as an ensemble model, is generally found to be more accurate than any of the individual models making up the ensemble (Dietterich, 2002).

Studies using Random Forest (Pal, 2005), which is a decision tree based ensemble learning algorithm, have reported a significantly high accuracy for crop classification. In this paper, we have examined the performance of Random Forest (RF) classifier for multi-crop classification using Sentinel 2 temporal data and compared the result with Naïve Bayes (NB) classifier. Next, we also discussed on the best bio window to produce an accurate in season crop map.

2. DATA AND METHODS

2.1 Study Area & Data

The classification was performed on a multi-temporal Sentinel 2 data during 22nd October 2017 and 1 March 2018 of *Lekoda* Gram Panchayat, *Ujjain* for Rabi season (Figure 1). Sentinel 2 data is available at a temporal resolution of 5 days and each

image has 13 spectral bands at different spatial resolution of 10m, 20m and 60m. All bands at 10m and 20m resolution were used in this study (Table 1). The analysis was done at 10m and therefore Band 5, Band 6, Band 7, Band 8a, Band 11 and Band 12 were first resampled to 10m through cubic convolution method and then layer stacked with the remaining four 10m bands.

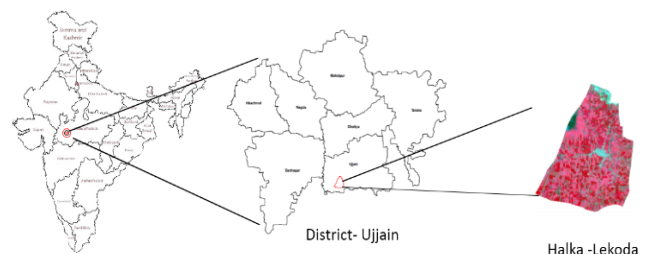


Figure 1. Study area selected in this study.

Sentinel 2 bands used in this study	Spatial Resolution
Band 2 – Blue	10
Band 3 – Green	10
Band 4 – Red	10
Band 5 – Red edge	20
Band 6 – Red edge	20
Band 7 – Red edge	20
Band 8 – NIR	10
Band 8A – Narrow NIR	20
Band 11 – SWIR	20
Band 12 - SWIR	20

Table1. Sentinel 2 bands used in this study

Ground survey of the study area was conducted during last week of December 2017 and various field information like crop

* Corresponding author

type, seeding date, crop health, were collected for 132 locations. . Wheat and Gram were the major crops in the region whereas other crops Garlic and Potato were also present but in small number. Staggered sowing was prevalent in wheat and two different types of Gram i.e. Desi and Hybrid were also present. To account for staggered sowing in wheat, 3 sub-class i.e. early sown wheat, mid sown wheat and late sown wheat were created. These sub classes were strictly derived from the field information. Early sown class includes wheat sown during September, mid sown class includes wheat sown during October and late sown class includes wheat sown during November and December (Figure 2). Altogether, 8 different classes were created from ground data points.



Figure2. Wheat sub classes to account for staggered sowing

Considering the complexity of having 8 classes in ground truth, sufficient data points in each class was required to build an efficient model for classification. The size of ground data points was increased with an assumption that the ground area within a buffer region of radius 10m around each observation had a same class type. Ground truth data and its revised number after buffer operation are presented in Table2.

Class Type	Number of Ground Observations	Revised number of Ground Observations
Early Wheat	6	54
Mid Wheat	33	297
Late Wheat	8	72
Desi Gram	29	261
Hybrid Gram	31	279
Garlic	12	108
Potato	2	18
Fallow Land/ Settlement/ Road	11	99

Table 2. Ground Truth data of the study area

2.2 Classification Method

The random forest classifier, which is a decision tree-based model, and Naïve Bayes classifier, which predicts the outcome based on Bayes theorem, were used in this study.

2.2.1 Random Forest: Random forest (RF) is a type of supervised ensemble learning algorithm, which combines the response of several decision trees to make more accurate prediction. Each decision tree is trained in isolation and grown in random subspaces of data by using Classification and Decision Tree (CART) methodology (Breiman et al., 1984).

RF creates many random sub samples (Bags) of the original dataset through Bagging. Bagging is a method to generate a training dataset by randomly drawing with replacement K examples, where K is the size of the original training set (Breiman 1996). For each sub sample, a decision tree is grown by random feature selection at each node of the tree .Final output is predicted based upon the majority vote of each of these decision trees. RF methodology and the functioning of an

individual tree are shown below in Figure 3 and Figure 4 respectively.

There are two user-defined parameters, which affects the performance of RF classifier. These parameters are (i) the number trees to be grown (N), and (ii) the number features to be selected at each node for best split (m).

Generally, the performance of RF gets better with increase in N. However, optimizing N becomes necessary since the improvement decreases as N increases i.e. the benefit of introducing an extra tree will be lower than the cost of computation For selection of m, Breiman (2002) criteria is used which says, if there are 'p' features in original dataset then only $m = \sqrt{p}$ randomly selected features at each node should be looked for best split.

In RF, the impact of overfitting an individual tree is less and therefore each tree is allowed to grow deep without pruning. This strategy makes RF unexcelled in accuracy when compared to other existing algorithms including discriminant analysis, support vector machines and neural networks (Liaw and Wiener, 2002).

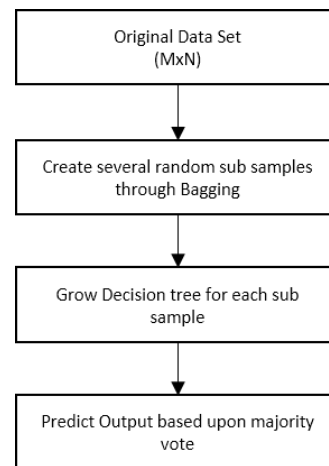


Figure 3. Random Forest Methodology

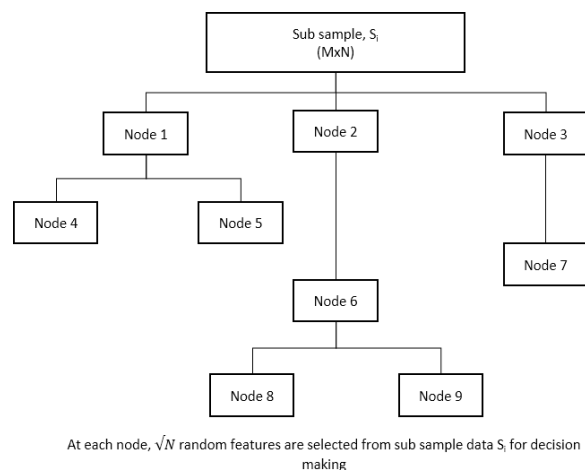


Figure 4. Functioning of an individual tree in RF

2.2.2 Naïve Bayes Classifier: The naive Bayes (NB) classifier comes under a family of probabilistic classifiers that predicts the outcome based on Bayes theorem with an

assumption that predictors are independent of each other (Domingos and Pazzani, 1997). The conditional independence among features may not be true for real world problems. However, Domingos *et. al*, 1996, has shown that NB performs well in practice even when this assumption is violated.

Let $X \{x_1, x_2, x_3, \dots, x_n\}$ denote the predictor variable and $C \{C_1, C_2, C_3\}$ be the response variable then the probability that X belongs to class $C \{C_1, C_2, C_3\}$ is given by:

$$P(C|X) = \frac{P(C)P(X|C)}{P(X)}$$

Where $P(C|X)$ is the probability that X belongs to class C , $P(C)$ is the prior probability of class C , $P(X|C)$ is the likelihood of the occurrence of X given that it belongs to class C and $P(X)$ is the probability of event X .

Assuming that all the predictors are conditionally independent of each other, eqn1 can be reduced to

$$P(C|X) = \frac{P(C) \prod_{i=1}^n P(X_i|C)}{P(X)}$$

NB classifier produces output based upon $P(C|X)$. For each class, NB computes $P(C|X)$ assign class C to X where $P(C|X)$ is maximum.

2.3 Model Training

There were 16 cloud free Sentinel 2 scenes available during 22 October 2017 and 1 March 2018. From each image, we selected ten features to build an ensemble learning classifier (Table1). In this study, we have accessed the performance RF and NB on multi-temporal satellite data for crop classification.

Classification performance on using multi-temporal satellite data single day satellite pass was compared as below;

Two separate categories of classifiers i.e. Category1 and Category2 were built. Classifiers under Category1 classify an image $I_i \{i = 1, 2, \dots, 16\}$, by using feature information of only single day, $D_i \{i = 1, 2, \dots, 16\}$, whereas classifiers under Category2 classify an image I_i by using feature information of all the available temporal data till D_i . For example, for D_5 , 10 features corresponding to I_5 will be used by Category1 classifiers and 50 features corresponding to I_1, I_2, I_3, I_4 and I_5 will be used by Category2 classifiers.

In subsequent sections, we have used the nomenclature, $RFC1_i$ and $NBC1_i$ to denote RF and NB classifiers under Category1 on day D_i , and $RFC2_i$ and $NBC2_i$ to denote RF and NB classifiers under Category2 on day D_i .

In this analysis, we have used 65% of ground data (774 data points) for training and 35% of ground data (413) to test the classifier. Number of training and test data for each class is shown below in Table 4.

Class Type	Number of training data	Number of test data
Early Wheat (C1)	35	19
Mid Wheat (C2)	194	103
Late Wheat (C3)	47	25
Desi Gram (C4)	170	91
Hybrid Gram (C5)	181	98
Garlic (C6)	71	37
Potato (C7)	12	6
Fallow Land/ Settlement/ Road (C7)	64	34

Table 3: Number of ground data used for model training and validation.

3. RESULT AND DISCUSSION

Classification accuracy of RF method depends on number of m and N . For $RFC1_i$'s, m was fixed at 4 and criterion suggested by Breiman (2002) was used to compute m for $RFC2_i$'s. We have used $RFC1_6$ to optimize N since its Kappa coefficient, κ , was highest among all $RFC1_i$'s (Figure 5). Optimum value of N was found to be 127 (Figure 6).

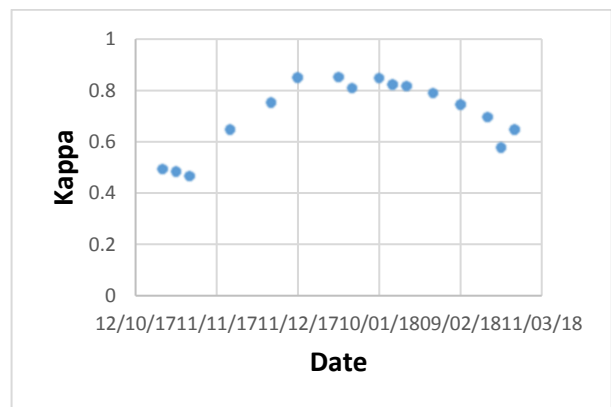


Figure 5. Kappa coefficient against each day for $RFC1$ classifier.

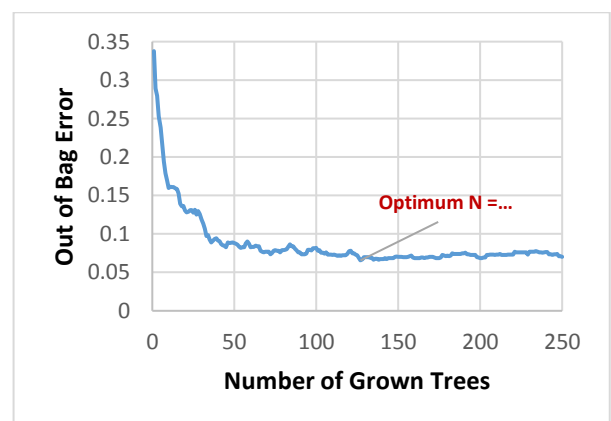


Figure 6. Progression of Out of Bag Error (OOBE) as the number of grown trees increases for classifier $RFC1_6$ which corresponds to D_6 i.e. Sentinel 2 data available on 12 December 2017.

Day	Kappa Coefficient			
	Category1		Category2	
	RFC1 _i	NBC1 _i	RFC2 _i	NBC2 _i
D ₁	0.49	0.27	0.49	0.27
D ₂	0.48	0.20	0.67	0.29
D ₃	0.46	0.30	0.78	0.39
D ₄	0.64	0.27	0.88	0.45
D ₅	0.75	0.43	0.90	0.63
D ₆	0.84	0.55	0.90	0.72
D ₇	0.84	0.68	0.93	0.77
D ₈	0.80	0.69	0.93	0.78
D ₉	0.80	0.75	0.93	0.79
D ₁₀	0.82	0.73	0.94	0.79
D ₁₁	0.81	0.61	0.93	0.8
D ₁₂	0.78	0.52	0.94	0.8
D ₁₃	0.74	0.47	0.95	0.81
D ₁₄	0.69	0.38	0.95	0.82
D ₁₅	0.57	0.40	0.95	0.81
D ₁₆	0.64	0.26	0.96	0.82

Table 5: Kappa Coefficients of RF and NB classifiers under Category1 and Category2.

During parameter optimization, misclassification among some classes was observed and therefore a cost matrix was introduced to improve classification accuracy (Table 4).

Classes	C1	C2	C3	C4	C5	C6	C7	C8
C1	0	1	1	20	1	1	1	1
C2	1	0	1	10	10	10	10	10
C3	1	1	0	10	10	10	10	10
C4	1	1	1	0	10	1	1	1
C5	1	1	1	5	0	1	1	1
C6	1	5	1	2	1	0	1	1
C7	1	5	1	2	2	1	0	1
C8	1	1	1	1	1	1	1	0

Table 4. An 8x8-cost matrix to account for error due to misclassification

Comparison between Category1 and Category2 classifier is shown in Table 6. Performance of Category2, which does classification based on multi-temporal Sentinel 2 data, was found to be better than the performance of Category1. Accuracy level of RF classifier was significantly higher than accuracy level of NB classifier in both the categories (Figure 5).

We found that the kappa coefficient of RFC1_i's during December to January was higher than 0.8 (Table 5). It so happen because during this period all the crops would attain a minimal above ground biomass than can be captured from Sentinel 2 spectral bands. The bio window during December to January would be the optimum time to generate in season crop map. However, the accuracy level of crop classification can be increased on using multi temporal data due to its high dimensionality and extra feature information (Figure 6). Kappa coefficient as of D₈ was 0.8 for RFC1 and 0.93 for RFC2.

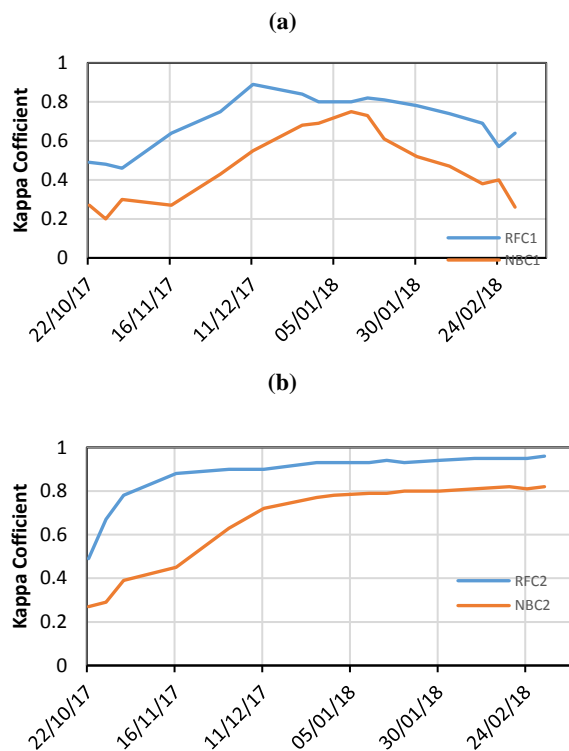


Figure 6: (a) Performance of RF and NB while using single day satellite data for model building, (b) Performance of RF and NB while using temporal data for model building.

RFC2₈ i.e. RF classifier in Category2 as of 31 December 2017 was deployed on the study area and an overall accuracy of 93% was achieved. Classified image of the study area is shown in Figure 7 and its corresponding confusion matrix and accuracy assessment is provided in Table6.

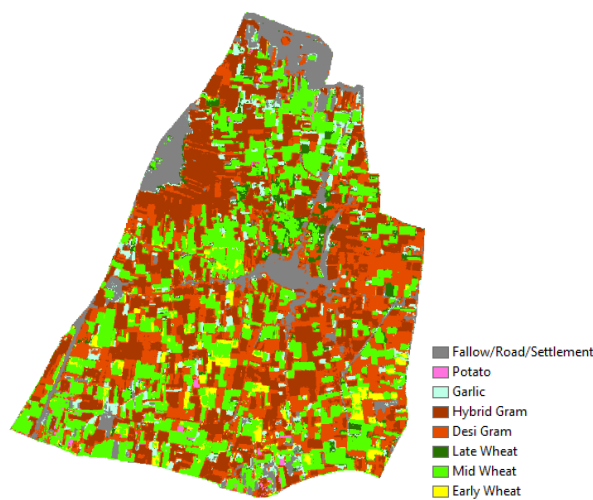


Figure 7: Classified Image of the study area as of 31 December 2017 using multi temporal satellite data and RF classifier.

(a)

Class	Producers Accuracy (%)	Users Accuracy (%)
C1	100	78.94
C2	95.13	96.11
C3	79.16	76.00
C4	90.52	94.50
C5	96.93	96.93
C6	100	94.59
C7	100	100
C8	100	100
Kappa = 93%		

(b)

Class	C1	C2	C3	C4	C5	C6	C7	C8
C1	15	0	1	3	0	0	0	0
C2	0	99	4	0	0	0	0	0
C3	0	3	19	3	0	0	0	0
C4	0	2	0	86	3	0	0	0
C5	0	0	0	3	95	0	0	0
C6	0	2	0	0	0	35	0	0
C7	0	0	0	0	0	0	6	0
C8	0	0	0	0	0	0	0	3
								5

Table 6: (a) Confusion Matrix corresponding to crop Classification done on the study area as of 31 December 2017 using multi temporal satellite data and RF classifier, and (b) Accuracy Matrix associated with it

4 CONCLUSION

In this study, we found that Random Forest classifier performed better than Naïve Bayes classifier. It is because Naïve Bayes assumes the conditional independence of the predictor variable whereas Random Forest does not make any assumption on predictor variables. The study concluded that the best bio-window for crop classification using single date satellite data would be during December to early January. The Kappa

coefficient was found to vary between 0.8 and 0.9 during this period. However, an increase of 12% in kappa coefficient was observed on using multi temporal Sentinel 2 data for the same period. It was also found that Random forest model could handle high resolution, multi-dimensional and multi temporal data set efficiently to produce accurate in season crop map.

REFERENCE

Bahram, Salehi, Bahram Daneshfar, Andrew M. Davidson. (2017) Accurate crop-type classification using multi-temporal optical and multi-polarization SAR data in an object-based image analysis framework. *International Journal of Remote Sensing* 38:14, pp. 4130-4155.

Breiman L. (2002) - Manual on Setting Up, Using, And Understanding Random Forests V3.1.

Breiman, L. Bagging predictors. *Machine Learning*, 24(2): 123-140, 1996.

Breiman, L., Friedman, J., Stone, C., and Olshen, R. *Classification and Regression Trees*. CRC Press LLC, 1984.

Dietterich, T.G., 2002, *Ensemble Learning*. The handbook of brain theory and neural networks, M.A. Arbib (Ed.) (Cambridge, MA: The MIT Press).

Domingos, P. and Pazzani, M. (1996). Beyond independence: Conditions for the optimality of the simple Bayesian classifier. *Proceedings of the 13th International Conference on Machine Learning*, Bari, Italy, pp. 105–112.

Domingos, P. and Pazzani, M. (1997). On the optimality of the

Liaw A., Wiener M. (2002) - *Classification and Regression by Random Forest*. *R News*, 2 /3, December

M. Pal (2005) Random forest classifier for remote sensing classification, *International Journal of Remote Sensing*, 26:1, pp. 217-222, DOI: [10.1080/01431160412331269698](https://doi.org/10.1080/01431160412331269698)

Simple Bayesian classifier under zero-one loss, *Machine Learning* (29), pp. 103–130.