Journal of AgriSearch, 8 (1): 21-25
An Open Access International Peer Reviewed Quarterly

ISSN : 2348-8808 (Print), 2348-8867 (Online)
https://doi.org/10.21921/jas.v8i01.19558

Traits preferred by birds to attack on standing wheat crop under temperate condition and its prediction through random forest model

ANIL KUMAR SINGH[1*], LAL SINGH[2], YASIN JK[3], ROHAN KUMAR RAMAN[1],PK SUNDRAM[1], PAWAN JEET[1], KIRTI SAURABH[1], VED PRAKASH, ASHUTOSH UPADHYAYA[1] AND UJJWAL KUMAR[1]

ABSTRACT

Wheat is an important cool season cereal crop of India, production and quality get compromised due to biotic and abiotic stresses. Wheat is susceptible to a variety of diseases and insect pests, including birds. Birds attack on standing crop are not very often and extent and pattern of damage varies considerably, however, its management practices is generally not taken in to account. The bird preferred to attack wheat crops while they were in the reproductive stage, especially during milking to dough stage. The data collected over the course of a two-year field study in the Kashmir valley at Research Farm Srinagar (Jammu & Kashmir) has been subjected to Random Forest model to determine whether there is a trait-specific specific preference for attack if any. Random forest (RF) model identified five major features out of 13 selected features of wheat germplasm with seventy percent accuracy for parrot birds'preferredto attack on wheat germplasm. RF Model suggests that tallerheight, having long spikes and a greater number of ears, along with more biomass and grain yields, could be possible traits that bird pest targetduring the reproductive stage. It was inferenced that the average performance of the random forest model demonstrated with model accuracy (71%), sensitivity (0.77), kappa 0.41 and receiver-operating-characteristic (ROC-0.74) and showed good fit for classification of parrot attack on wheat germplasm.

KEYWORDS

Bird attack, Kashmir, Random forest model, Traits, Wheat

ARTICLE INFO

Received on	:	14/07/2020
Accepted on	:	10/01/2021
Published online	:	19/03/2021

INTRODUCTION

Wheat play a major role in agricultural produce to satisfy the global food demand (Chand *et al.*, 2017). In India, wheat is second most important cereal crop after rice, where majority of the population fulfill their nutrition and calorie intake through cereal based production system (Singh *et al.*, 2017). By nature, wheat crop is a basically temperate crop and has potential in terms of production and productivity (Meena *et al.*, 2016 and Singh *et al.*, 2017). Kashmir valley conditions in India comes under geographically temperate zone but still very rare study were found on wheat production and potential for this region(Singh *et al.*, 2015).Agriculture production system severely affected due to several unpredictable multi-faceted challenges like potential factor, limiting factor and reducing factor(Meena *et al.*, 2013 and Singh, 2016). Productivity can be improved through efficient management of limiting and reducing factors like water, nutrients and pest's *viz*., disease insects and other pests(Singh *et al.*, 2017). Above mentioned factors and their management are very common in nature; however some of the uncommon pests like birds pest management are not so common. Though there are management practices to deter rodent and other ruminant attacks, studies and analyses of the magnitude and pattern of harm caused by birds' pest are scarce. While a few studies on birds attack in India's plains have been performed on a small scale, there is a lack of knowledge on bird attack and damage in India's temperate valleys..Since, study on crop damage and consequently yield losses by bird attacks are not so common, hence data and literature on this aspect is also scanty. Secondly, the environmental parameters interaction with biotic community are also generally non-normal and complex in nature. The complex interactions of environmental parameters that vary in time and space put big challenges in front of researchers to understand and find the exact relationship (Prasad *et al*., 2006). The inference drawn from the complex data through traditional statistical analysis provides invalid inference due to lack of underlying assumptions of traditional statistical model. Machine learning modeling is an interdisciplinary nonparametric computational technique in ecological informatics that is very much efficient in coping with highly complex and non-linear problems (Lehikoinen *et al.*, 2019). Recently many computational statistical methods have emerged from machine learning algorithms that are capable of handling ecological data mining. Random forest is one such method of machine learning that is used for both classification and regression analysis for ecological data (Prasad *et al*., 2006). This method RF can also handle the data easily when the number of features is much larger than the number of samples. Feature selection technique is a tool of machine learning algorithm that minimizes redundancy and maximizes relevancy for classification problems. Here a target values of

[1]ICAR-Research Complex for Eastern Region, Patna, Bihar, India
[2]Sher-e-Kashmir University of Agriculture Sciences and Technology of Kashmir, Srinagar (J&K) India
[3]ICAR-National Bureau of Plant Genetic Resources, New Delhi, India
*Corresponding author email : anil.icarpat@gmail.com

data set is predicted using an identified potential set of data. Random forest were applied in different discipline like agriculture, fisheries, medical science and engineering for important feature selection (Shanmuga and Abinaya, 2018; Koteswara *et al*., 2019).

Hence this study is coined to identify the important features of wheat germplasm attack by the bird (parrot) in the standing wheat field crop in premature stage in the temperate zone Kashmir valley, India using machine learning Random Forest modeling.

MATERIALS AND METHODS

Observations and data recording

Field experiment was conducted at National Bureau of Plant Genetic Resources, Regional Station, and Srinagar, Jammu and Kashmir, India. The location of the study site was at 34^0 5 ' N latitude and 74^0 8 ' E longitude and altitude 1605 m. Total 119 number of wheats germplasms were evaluated in winter (*rabi*) seasons for two consecutive years to draw any valid inference (Table 1). Augmented randomized complete block design (Johnson *et al*., 1955) has been applied in this field experimentation. Each wheat genotypes / germplasm was sown in 4 lines of 3-meter row, keeping row distance interval of 23 cm. During first- and second-year crop was sown on 5th October and 7th October and harvested on 21st June 2001 and 27th June, respectively. All the agronomic measures were taken in to account for raising and maintaining a healthy crop.The observation were taken from pre earmarked 1 meter row length randomly at two places in each plot through augmentation design. Data were recorded on biometric *i.e.,* agronomic as well as yield attributes and seed yield of wheat, as plant height (cm), days to heading (number), days to anthesis (number), days taken to maturity (number), no. of ear (number), spike length (cm), number of spikes, grains per spike, test weight (g) of seed, dry weight per shoot (gm), grain weight per shoot, biomass yield (g/plant), grain yield (g) and harvest index (HI) of every wheat germplasm. Data were also recorded on parrots (birds) attacked on some of the standing wheat germplasm and damaged germplasmsa nd their preference/choice of particular germplasm, and data were also taken in binary form (yes or no) for attacked or not attacked germplasm.

Random Forest (RF) Model development

Random Forest (RF) is a supervised and tree-based ensemble machine learning approach that works on mixture of decision trees;{T1(x),...,TB(X),} where X={x1,...,xp} is a *p*-dimensional vector of variables. Each split of the tree follows the concept of bootstrap aggregation (*e.g.,* subsampling input samples with replacement) and random subspace method (e.g., subsampling the variables without replacement). The ensemble produces B outputs {Y1=T1(X),...,YB+TB(X)} where

Table 1: Wheat genotypes evaluated at NBPGR Regional Station Srinagar (Jammu and Kashmir) India

Sr. No	Accession (IC No.)	Sr. No	Accession (IC No.)	Sr. No	Accession (IC No.)	Sr. No	Accession (IC No.)	Sr. No	Accession (IC No.)
1	82411	25	73571	49	29065	73	79520	97	78777
2	82424	26	73574	50	31616	74	79703	98	78779
3	82426	27	73575	51	35100	75	78976	99	78780
4	82435	28	73577	52	35130	76	78708	100	78787
5	82440	29	73579	53	35145	77	78111	101	78788
6	82442	30	73580	54	35152	78	78713	102	78793
7	82452	31	73884	55	35153	79	78722	103	78799
8	82453	32	73599	56	35155	80	78725	104	78810
9	82458	33	73645	57	35185	81	78726	105	78811
10	82526	34	76854	58	55716	82	78735	106	78813
11	82620	35	78700	59	57586	83	78737	107	78819
12	82796	36	78703	60	78994	84	78739	108	78821
13	82803	37	78705	61	79002	85	78741	109	78830
14	104652	38	78707	62	79007	86	78743	110	78835
15	104657	39	28532	63	79008	87	78744	111	78759
16	104698	40	28696	64	79105	88	78746	112	78878
17	63947	41	28835	65	79027	89	78748	113	78942
18	66518	42	28852	66	79031	90	78752	114	78947
19	73209	43	28871	67	79034	91	78754	115	78950
20	73214	44	28904	68	79047	92	78758	116	78956
21	73491	45	28908	69	79048	93	78765	117	78964
22	73493	46	28959	70	79050	94	78766	118	78968
23	73495	47	29023	71	79075	95	78768	119	78975
24	73498	48	29062	72	7983	96	78771		

Weekly average maximum temperature ranged between 7.81 to 29.41^0C and minimum -5.5 to 13.53^0C during the cropping season both the years (Fig. 1).

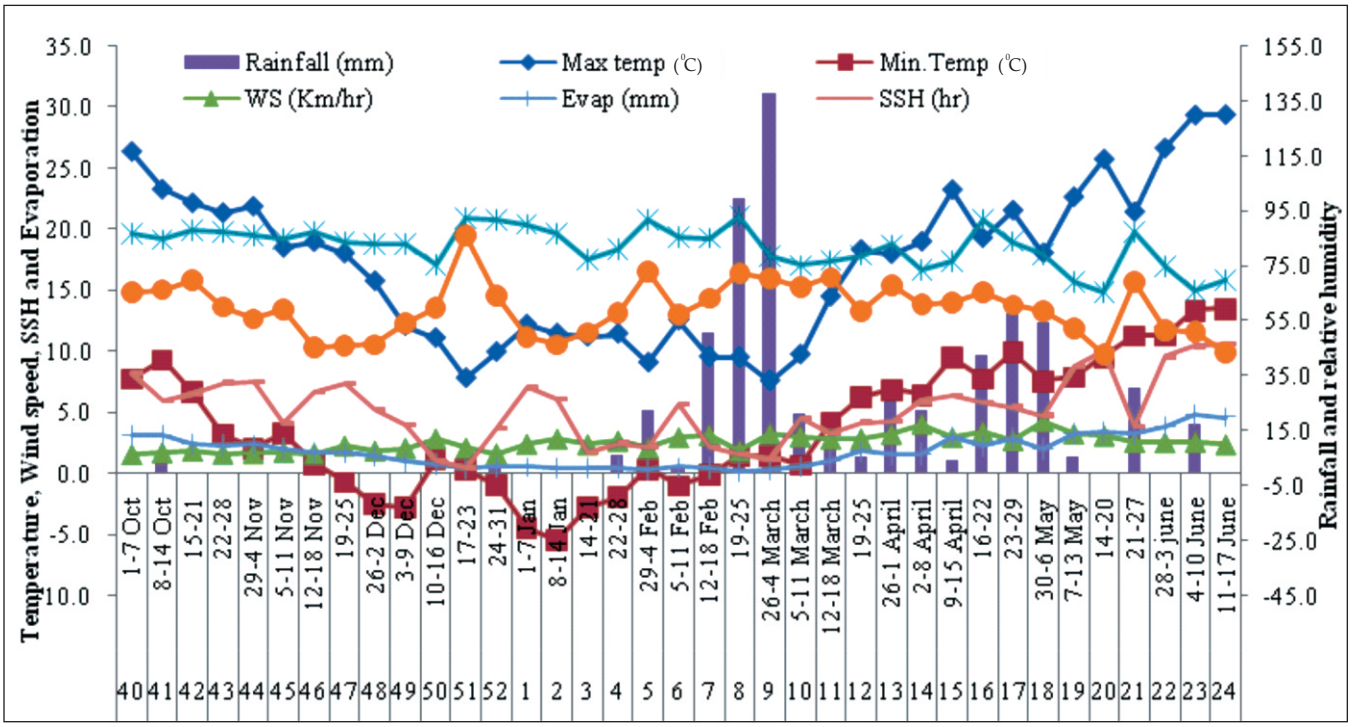

Fig. 1: Weekly average temperature, wind speed, relative humidity, sunshine hours evaporation and total rainfall during crop growth period of *rabi* season 2002-03 at experimental site

YB,b=1,...,B is the predicted weight by the bth tree. The class Y is predicted by the majority of trees (Breiman, 2001). RF has advantage that it does not over-fit and has robustness to noise and irrelevant features to produce good predictors. Here, R package "random Forest" used for prediction modeling. Here RF model is developed with the following steps. At first total data sets are divided into a training set (80% of the samples) and validation set (20% of the samples). The training set is used for model buildings for classification and validation set is used to assess the prediction accuracy of the model through random sample selection. The performance of the model is evaluated through four model evaluation criteria based on confusion matrices like (a) classification accuracy e.g., percentage of correctly classification instances (CCI) (b) sensitivity of the model (ability to accurately predict the attack of parrot on different wheat germplasm and (c) Cohen's kappa coefficient (*k*), a reliable performance measure (Tirelli *et al*., 2009) and (d) the area under the receiver-operating-characteristic (ROC) curve. In the area under the ROC curve, a value of 0.7, 0.8 and 0.9 indicates satisfactory, good and very good discrimination respectively. Here 10-fold cross validation is taken into consideration for the average predictive performance of the RF model.

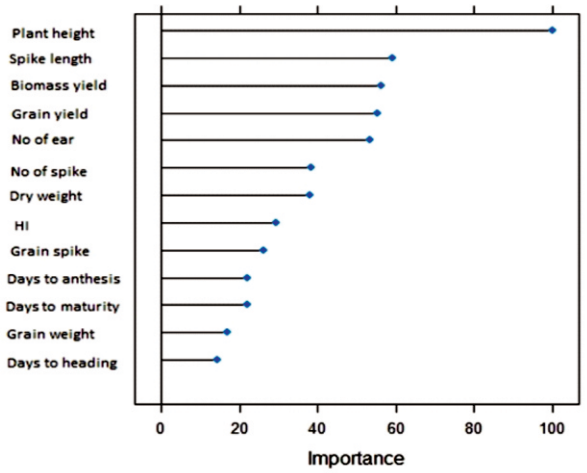

Fig.2: Important wheat germplasm features identified by Random Forest modelfor parrot attack at Kashmir, India

RESULTS AND DISCUSSION

Biometric observation

Table 2 indicates the average and standard deviation of five randomly chosen wheat plants from 119 genotypes.Height of plant is one of the important agronomic traits which influence the overall plant geometry, their growth, development and final output in terms of economic yield. Mean plant height of all 119 accessions was observed to be 78.35 cm with SD 10.2. This indicate rich diversity was present in wheat germplasm under study. Maximum variation was recorded in yield attributes in terms of standard deviation (281.95) in case of biomass yield (g/plant) followed by grain yield (158.50 g) and number of ear (70.05). However, minimumvariation was recorded in case of HI (0.04), dry wt. / shoot (0.15) and grain wt. /shoot (0.18).

The bird attack affected 39 wheat germplasm out of 119 wheat germplasm analysed in field conditions, while 75 wheat germplasm is unaffected. The Random Forest model was used to identify affected and nonaffected wheat germplasm, as well as the 13 associated selected parameters.The average

Table 2: Mean and standard deviation of the 119 wheat genotype data for the period 2001-03 in Kashmir, India

Characters	N	Mean	Std. Dev.
Plant Height (cm)	119	78.35	10.20
Days to heading	119	202.46	2.47
Days to anthesis	119	208.46	2.47
Days to maturity	119	239.51	2.16
No of Ear	119	158.45	70.05
Spike length (cm)	119	7.78	1.19
No of Spike	119	15.70	10.53
Grains/spike	119	43.13	5.24
Dry Wt./ Shoot	119	3.84	0.15
Grain Wt./Shoot	119	2.08	0.18
Biomass Yield (g/plant)	119	614.26	281.95
Grain Yield (g)	119	335.46	158.50
HI	119	0.54	0.04

performance of the random forest model exhibited model accuracy (71%), sensitivity (0.77), kappa 0.41 and ROC (0.74) and showed good fit for classification of parrot attack on wheat germplasm. These model validation parameters variation was plotted in **Fig.3**.

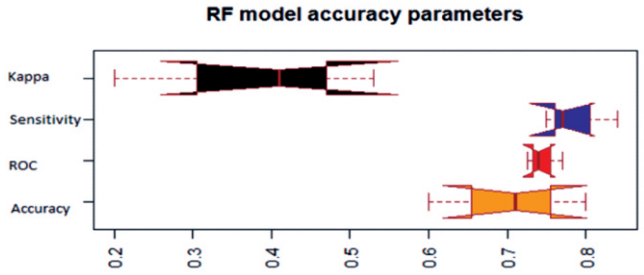

Fig. 3: Random forest model evaluation parameters (kappa, sensitivity, ROC and accuracy) for classification of parrot attacked on wheat germplasm in Kashmir, India

Distinctive traits favored birds attack

A random forest model for feature selection for parrots attacking wheat germplasms revealed that five parameters out of thirteen play a significant role in attracting parrots in the field. Careful observation of traits recorded during the course of study/ field experimentation indicates that unique characters such as plant height, spike length, biomass yield, grain yield and number of ears are the unique features that invites / favors bird attraction to attack and damage the crop (**Fig.3**). These five identified features have higher value in magnitude that favors the bird to attack on these crops (**Table 3**). It is clearly evident from the data that bird prefers to attack on those wheat plant which was comparatively taller among the tested genotypes/ germplasm, probably, due to its tallness of plants, this might be very easy to identify plant attack. Second crucial plant traits which birds prefers was number of ears per running meter. If the genotypes have more ears the chances are more to get attacked by the birds. Spike length has direct and positive relation with bird attack and damage, in other words, genotypes with longer ear are more likely to be attacked and harmed. Similarly, the greater the biomass production, the more likely the plant is to be targeted by birds; this may be attributed to the plant's increased strength, which allows it to withstand the weight of the bird without breaking and with less bending. Similarly grain yield (g) has positive influence on bird attack on wheat germplasm. Wheat germplasm with a higher wheat grain yield has a higher risk of being targeted by birds than wheat germplasm with a lower grain yield. (**Table 3**).

Table 3: Mean and standard deviation of the affected and not affected wheat germplasmby the parrot at Kashmir, India

Parameters	Affected (39 No.)		Nonaffected (75 No.)	
	Mean	Std Dev	Mean	Std. Dev
Plant Height (cm)	83.05	8.33	75.60	10.24
Days to heading	204.64	15.57	202.52	2.31
Days to anthesis	210.64	15.57	208.52	2.31
Days to maturity	239.45	2.42	239.55	2.02
No of Ear/ Meter	184.52	66.57	143.15	67.88
Spike length (cm)	7.93	1.01	7.69	1.28
No of Spike	14.91	1.41	16.17	13.23
Grains/spike	43.68	4.19	42.81	5.76
Dry Wt./ Shoot	48.16	2.10	48.76	2.12
Grain wt./shoot	3.88	0.12	3.81	0.16
Biomass Yield (g/plant)	719.48	265.66	552.53	274.43
Grain Yield (g)	387.75	144.78	304.77	159.06
HI	2.10	0.12	2.08	0.20

Conclusion

Wheat is one of the major crop attacked by birds under Kashmir valley condition. Random forest model, a machine learning tool is useful for classification and feature identification. Model, suggest that birds prefer to attacks plants which are taller, having long and a greater number of ears, along with more biomass yield and grain yield. The accuracy of prediction made by random forest model (RFM) in this study was (71%).

Acknowledgment

The authorsacknowledges the support of National Bureau of Plant Genetic Resources, New Delhi, forproviding all assistance and wheat germplasms for evaluation under Kashmir conditions.

REFERENCES

Breiman L. 2001. Random forests. *Machine Learning* **45**:15–32.

Chand Subash, Singh AK and Kumar R.2017. Effects of tillage practices on productivity of wheat under the Indo-Gangetic Plains. *Journal of AgriSearch* 4 (3):209-2011.

Johnson HW, Robinson HF and Comstock RE 1955. Estimates of genetic and environmental variability in soybeans. *Agronomy journal* **47**(7): 314–318.

Koteswara C, Chinnababu M and Sarangarm K. 2019. Classification of Diabetes using Random Forest with Feature Selection

Algorithm.*International Journal of Innovative Technology and Exploring Engineering* **9**(1): 1295-99.

Lehikoinen A, Olsson J, Bergström L, Bergström U, Bryhn A, Fredriksson R and Uusitalo L. 2019. Evaluating complex relationships between ecological indicators and environmental factors in the Baltic Sea: A machine learning approach. *Ecological indicators***101**:117-125.

Meena BL, Singh AK, Phogat BS and Sharma HB. 2013. Effects of nutrient management and planting systems on root phenology and grain yield of wheat. *Indian J. Agril. Sci.* **83** (6): 627-632.

Meena BL, SinghAK, Phogat BS and Sharma HB.2016.Improving wheat and soil productivity through integrated nutrient management (INM) and efficient planting system (EPS). *Journal of AgriSearch* **3**(3): 147-156.

Prasad AM, Iverson LR and Liaw A. 2006. Newer classification and regression tree techniques: bagging and random forests for ecological prediction. *Ecosystems* **9**:181–199.

Shanmuga PS and Abinaya M.2018. Feature Selection using Random Forest Technique for the prediction of pest attack in cotton crops.*International Journal of Pure and Applied Mathematics***118** (18): 2899-2903.

Singh AK,Pedapati A, Manibhushan, N Chandra N and Narayan B.2015. Extent and pattern of diversity in saffron germplasm of Indian Kashmir. *Bangladesh Journal of Botany***44** (4):635 642.

Singh AK, Singh AK, Kumar R, Prakash V, Sundaram PK and Yadav SK. 2017. Indian Cereals Saga: Standpoint and Way Forward. *Journal of AgriSearch* **4** (1): 1-10.

Singh AK.2016. Understanding weed biology to manage wheat field efficiently. *Indian Farming***66** (6): 17-21.

Tirelli T, Pozzi, L and Pessani D.2009. Use of different approaches to model presence/absence of Salmo marmoratus in Piedmont (Northwestern Italy), *Ecological Informatics***4**(4):234-242.

Citation:

Singh AK, Singh Lal, Yasin JK, Raman RR, Sundram PK , Pawan Jeet, Saurabh Kirti, Prakash Ved, Upadhyaya Ashutosh and Kumar Ujjwal. 2021. Plant traits preferred by birds to attack on standing wheat crop under temperate condition and its prediction through random forest model. *Journal of AgriSearch* **8**(1): 21-25