Available online at www.isas.org.in/jisas
Journal of the Indian Society of Agricultural Statistics 75(2) 2021 141–145

Evaluating Machine Learning Approaches for Prediction of Suitable Climatic Conditions for *Parthenium hysterophorus* (L.) in India

Abhishek Yadav[1], Yogita Gharde[2] and Sushil Kumar[2]

[1]*Indira Gandhi Krishi Vishwavidyalaya, Raipur*
[2]*ICAR-Directorate of Weed Research, Jabalpur*

Received 04 October 2021; Revised 02 November 2021; Accepted 10 November 2021

SUMMARY

Parthenium hysterophorus is considered as one of the worst weed of the world as it is difficult to control. Besides extensive documentation on the weed, the details of its current range with intensity level are not fully known in India. Further, Machine Learning algorithms are considered as promising approaches for modelling and predicting the distribution of different species. Hence, the present study aims in evaluating different machine learning approaches (MLA) to find out the best one in predicting the suitable climatic conditions for *Parthenium hysterophorus* in India. Climatic variables such as mean maximum temperature, mean minimum temperature, relative humidity and rainfall were used as independent variables for prediction. Total of 14 machine learning algorithms were included in the evaluation process. The best algorithm for the prediction was chosen using criteria like percentage of correctly classified instances, Root Mean Squared Error (RMSE), Mean Absolute Error (MAE) and Root Relative Squared Error (RRSE). Among all machine learning algorithms, J48 was found best for the prediction of suitable climate conditions for establishment of Parthenium in India. This study is helpful in knowing the possible infestation level of Parthenium in a place based on weather conditions which may further be helpful in planning the management strategies timely.

Keywords: Machine Learning algorithms, Modelling, Alien invasive weed, *Parthenium hysterophorus.*

1. INTRODUCTION

Weeds are considered as significant biotic constraints in agricultural production systems and pose harmful threats to agro-biodiversity and aquatic bodies. Problem of weeds is increasing day by day mainly due to the adoption of modern crop cultivation practices e.g. use of mono-cropping systems without legumes, intensive cultivation, disproportionate use of chemical fertilizers, development of herbicide resistance in weeds, invasion of alien weeds, consequence of climate change favouring more rigorous growth of weeds and risk associated with herbicide residue. In recent studies, total economic losses arising due to weeds in 10 major field crops in 18 states of India were estimated approximately US$11 billion (Gharde *et al.* 2018). This cost has posed major challenges for weed researchers and policy makers. Further, Alien Invasive Species (AISs) are the second biggest threat to biodiversity after habitat destruction (CBD, 1992) and are a major challenge to the economic well-being of the globe. Further, management of Alien Invasive Weeds (AIWs) depend upon the information about their projected distributional potential and relative spread in the current situation. This information is necessary for assessment of risk due to AIWs as well as for planning of suitable long-term management strategies. Prevention of an AIW's establishment and further future spread is known as a far more efficient and less costly than other management strategies such as eradication, containment and control which may be required when the species has established (Simberloff *et al.* 2013).

Parthenium hysterophorus (L.) (Asteraceae), commonly known as Parthenium is considered as one of the worst weeds of the world (Gharde *et al.* 2019). This weed is invasive in almost all countries around

Corresponding author: Yogita Gharde
E-mail address: yogitagharde@gmail.com

the globe, except Europe and some islands (Adkins and Shabbir, 2014). The weed has adapted in diverse environments having variable climatic conditions. It has been included in the list of Global Invasive Species Database (GISD, 2017) and is one of the most difficult weeds to control. In India also, it has invaded most of the states including areas with extreme climatic conditions (Kumar *et al.* 2008, Gnanavel 2013).The weed may have come into India from United States of America in a wheat food grain lot imported into Pune in 1956 (Rao, 1956; Vartak 1968) and subsequently spread throughout most of the areas of the country. It has resulted yield losses up to 40% in various crops and about 90% reduction in forage production (Gnanavel 2013). It has achieved the status of "worst weed" in India (Sushilkumar, 2014).

Further, use of machine learning algorithms are considered as promising approaches for modelling and predicting the distribution of different species (Elith *et al.*, 2006). Ahmad *et al.* (2019) used ensemble modelling using the ten statistical and machine learning algorithms for predicting current and future invasion of Parthenium with presence-absence data. Gharde *et al.* (2019) also used machine learning approaches to find out rules for predicting the distribution of *Zygogramma bicolorata* for the control of Parthenium.

Despites extensive documentation on the *Parthenium hysterophorus* abundance and distribution throughout the invaded continents, the details of its current range with variable intensity are not fully known in India. Therefore, present study was planned to evaluate different machine learning approaches for prediction of suitable climatic conditions of *Parthenium hysterophorus* in India using data on different level of intensity of infestation.

2. MATERIALS AND METHODS

2.1 Data collection

A survey was conducted between 2010 and 2019 to collect the data on occurrence of *Parthenium hysterophorus* throughout the India. Occurrence data available for 130 districts of India covering most of the states were considered for the study. The data collected is in different intensity such as negligible, moderate and high. Data collection sites are depicted in Fig. 1.

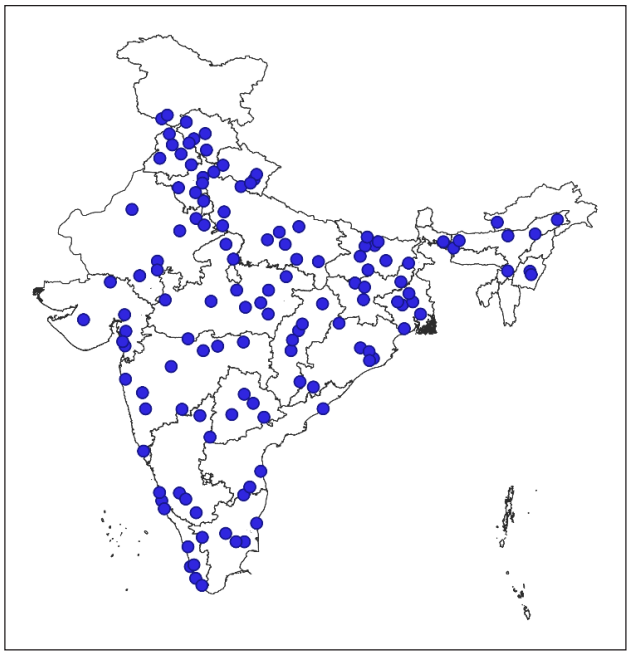

Fig. 1. Depiction of sites considered for the study

Data on climatic parameters were also collected from Indian Meteorological Department, Pune, India for the period from January, 2010 to December, 2019 (120 months). Climatic variables, namely mean maximum temperature (MMAX), mean minimum temperature (MMIN), relative humidity (RH) and rainfall (RF) were used as independent variables, whereas, occurrence level of Parthenium (*viz.* Negligible, Moderate, High) was considered as dependent variable.

2.2 Machine Learning algorithms for Modelling

Models were developed using different machine learning algorithms and they were compared for their performance in predicting the distribution of Parthenium occurrence. Among Machine Learning Approaches, different algorithms such as Random Forest, lazy.LWL, Bayes Net, logistic regression, Support Vector Machine, Decision stump, J48, Multi Layer Perception, REPT Tree, Hoeffding Tree, Logistic model tree, Simple logistic, Naive Bayes and Naive Bayes Multinomial were used. Decision tree was used to find the rules to predict the establishment. The decision tree is comprised of the internal nodes (different attributes as independent variables) and branches between the nodes (possible values that the attributes may have in the observed data), while the terminal nodes tell the final value of the dependent variable (Witten and Frank, 2005). Average value of four independent weather variables (MMAX, MMIN,

Table 1. Number of *Parthenium hysterophorus* infested sites in different states of India selected for the study

State	Number of observed sites
Andhra Pradesh	8
Arunachal Pradesh	1
Assam	6
Bihar	7
Chhattisgarh	6
Goa	1
Gujarat	6
Haryana	5
Himachal Pradesh	5
Jammu &Kashmir	2
Jharkhand	5
Karnataka	7
Kerala	6
Madhya Pradesh	10
Maharashtra	8
Manipur	3
Odisha	6
Punjab	5
Rajasthan	7
Tamil Nadu	5
Uttar Pradesh	10
Uttarakhand	5
West Bengal	6
Total	130

RH, RF) were considered as independent variables for this analysis. In the study, decision tree with different algorithms was formed using Weka 3.8.3 software.

2.3 Criteria for selection of algorithm

The best algorithm for the prediction of occurrence level of Parthenium in India was chosen using criteria such as Percentage of correctly and in-correctly classified instances among total dataset, Root Mean Square Error (RMSE), Mean Absolute Error (MAE) and Root Relative Squared Error (RRSE). Lesser value of these errors was considered the basis for selection of the model.

Root Mean Square Error (RMSE)

It is the standard deviation of the residuals (prediction errors). The RMSE tells us how concentrated the data is around the line of best fit. It is given by,

$$RMSE\left(\bar{X}\right)=\sqrt{E[\left(\bar{X}-\mu\right)^2]}\ =\ \sqrt{(\frac{\sigma}{\sqrt{n}})^2}\ =\sqrt{\frac{\sigma^2}{n}}$$

where σ^2 is the population variance and $\bar{X}=\frac{1}{n}\sum_{i=1}^{n}X_i$

X_i denote the observed response from i^{th} object and n total number of objects.

Mean Absolute Error (MAE)

It measures the average magnitude of the errors in a set of predictions, without considering their direction. It is the average over the test sample of the absolute differences between predicted and actual observations where all individual differences have equal weight. Lesser the value of MAE, better is the model. It is given by,

$$MAE=\frac{\sum_{i=1}^{n}|y_i-x_i|}{n}$$

where y_i is the predicted value and x_i is the true value and n total number of objects in the sample.

Root Relative Squared Error (RRSE)

Relative squared error takes the total squared error and normalizes it by dividing it with total squared error of the simple predictor. By taking the square root of the relative squared error, one reduces the error to the same dimensions as the quantity being predicted.

$$RRSE=\sqrt{\frac{\sum_{j=1}^{n}\left(P_{(ij)}-T_j\right)^2}{\sum_{j=1}^{n}\left(T_j-\bar{T}\right)^2}}$$

where, *P(ij)* is the value predicted by the individual model i for record j (out of n records); *Tj* is the target value for record j; and $\bar{T}$ is given by the formula:

$$\bar{\boldsymbol{T}}=\frac{1}{n}\sum_{j=1}^{n}T_j$$

3. RESULTS AND DISCUSSION

For the purpose, weather parameters such as mean maximum temperature (MMAX), mean minimum temperature (MMIN), rainfall (RF) and relative humidity (RH) were considered as independent variables and occurrence level of Parthenium (Negligible, Moderate and High) was considered as dependent variable. Summary about the values of the independent weather variables in the study are given in Table 2.

Table 2. Statistics of climatic parameters considered for study

Climatic parameter	Lowest value	Highest value
MMAX (ºC)	22.4	34.5
MMIN (ºC)	11.1	26.0
Rainfall (mm)	101	2348
Relative humidity (%)	30	78

Machine Learning approaches using different algorithms such as Bayes Net, logistic regression, Support Vector Machine, lazy.LWL, Decision stump, J48, Multi-Layer Perception, REPT Tree, Hoeffding Tree, Logistic model tree, Simple logistic, Naive Bayes and Naive Bayes Multinomial were used for model building in Weka. Results obtained from the analysis is presented in Table 3.

Table 3. Comparative results of different algorithms under different machine learning approaches

S. No.	Algorithm	Correctly classified data (%)	Incorrectly classified data (%)	MAE	RMSE	RRSE (%)
1.	J48	80	20	0.188	0.3066	71.81
2.	Lazy.LWL	75.38	24.62	0.2286	0.3387	79.33
3.	Multi-Layer Perception	73.85	26.15	0.2522	0.3549	83.12
4.	Bayes Net	72.31	27.69	0.2187	0.3649	85.47
6.	Logistic Regression	72.31	27.69	0.2619	0.3609	84.57
7.	Decision Stump	72.31	27.69	0.2872	0.379	88.75
8.	REPT Tree	72.31	27.69	0.2872	0.379	88.75
9.	Hoeffding Tree	71.54	28.46	0.2154	0.3873	90.71
10.	Logistic Model Tree	71.54	28.46	0.266	0.3615	84.66
11.	Simple Logistic	71.54	28.46	0.266	0.3615	84.66
12.	Support Vector Machine	71.54	28.46	0.3111	0.4037	94.50
13.	Naive Bayes	70.77	29.23	0.2154	0.3873	90.72
14.	Naive Bayes Multinomial	63.08	36.92	0.2541	0.4596	107.63

Based on the criteria viz. correctly classified instances, MAE, RMSE and RRSE; J48 algorithm was selected for model building and for obtaining decision tree as it classified 80% of data correctly and also has less error rate (MAE-0.188, RMSE-0.3066, RRSE-71.81%) as compared to other algorithms. Next best performing algorithm was Lazy.LWL (correctly classified instances-75.4, MAE-0.2286, RMSE-0.3387, RRSE-79.33%) followed by Multi-Layer Perception (correctly classified instances-73.9, MAE-0.2522, RMSE-0.3549, RRSE-83.12%). Further decision tree was obtained using J48 algorithm and rules for predicting the suitable conditions or weather parameters values for occurrences of *Parthenium hysterophorus* were drawn from the tree. Decision tree obtained using J48 algorithm is presented in Fig 2.

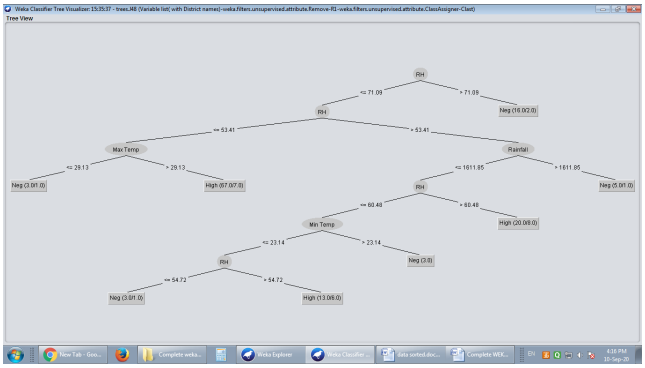

Fig. 2. Decision tree diagram obtained using J48 algorithm

Table 4. Rules obtained through J48 algorithm for prediction of occurrence level of *Parthenium hysterophorus*

S. No.	Rules	Predicated class or level of occurrence
1.	RH (71-78%)	Negligible
2.	RH (53 – 71%), Rainfall (1612-2348 mm)	Negligible
3.	RH (60 – 71%), Rainfall (<1612 mm)	High
4.	RH (55-60%), MMIN (11-23°C), Rainfall (<1612 mm)	High

Other instances, that do not follow the above mentioned rules, will fall in moderate class of occurrence level of Parthenium. According to these rules, area with high humidity (>71 %) and high rainfall (>1612 mm) are not suitable for Parthenium occurrence. Hence, coastal areas may not be suitable for Parthenium establishment. However, climatic conditions with mean minimum temperature (MMIN) between 11-23°C along with relative humidity 55-60 % and rainfall <1612 mm was found to be highly suitable for the occurrence of Parthenium.

4. CONCLUSION

Alien invasive weeds such as *Parthenium hysterophorus* has devastating effect on native plant species, causing decline or even extinctions of some of them, and negatively affecting environment and ecosystems. The present study evaluated machine learning algorithms for modelling and predicting

the distribution of *P. hysterophorus* with weather parameters as independent variables. Among all, J48 algorithm was chosen best for obtaining decision tree showing lesser error rate as compared to others. The algorithm shows that the areas with high humidity and high rainfall are not suitable for the establishment of *P. hysterophorus*.

This study will be helpful in identifying the possible infestation level of Parthenium in a place based on weather data which may further be helpful in planning the management strategies timely. Further, selected machine learning algorithm may be used to predict the future expansion of *Parthenium hysterophorus* in climate change scenarios.

REFERENCES

Adkins S.W. and Shabbir A. (2014). Biology, ecology and management of the invasive Parthenium weed (*Parthenium hysterophorus* L.). *Pest Management Science* 70:1023–1029.DOI 10.1002/ps.3708

Ahmad R., Khuroo A.A., Hamid M., Charles B. and Rashid I. (2019). Predicting invasion potential and niche dynamics of *Parthenium hysterophorus* (Congress grass) in India under projected climate change. *Biodiversity and Conservation* 28(8-9):2319-2344. https://doi.org/10.1007/s10531-019-01775-y

CBD, 1992. The Convention on Biological Diversity of 5 June 1992.

Elith J, Graham CH, Anderson RP *et al.* (2006) Novel methods improve prediction of species distributions from occurrence data. *Ecography,* **29**, 129-151.

Gharde Y., Singh P.K., Dubey R.P. and Gupta P.K. (2018). Assessment of yield and economic losses in agriculture due to weeds in India. *Crop Protection* 107:12–18. https://doi.org/10.1016/j.cropro.2018.01.007

Gharde Y., Kumar S. and Sharma A.R. (2019). Exploring models to predict the establishment of leaf-feeding beetle *Zygogramma bicolorata* (Coleoptera: Chrysomelidae) for the management of *Parthenium hysterophorus* (Asteraceae: Heliantheae) in India. *Crop Protection* 122:57-62.https://doi.org/10.1016/j.cropro.2019.04.014

Global Invasive Species Database. (2017). Species Profile: *Parthenium Hysterophorus*. Downloaded from. http://www.iucngisd.org/gisd/speciesname/on 05-07-2017.

Gnanavel I (2013). *Parthenium hysterophorus*: a major threat to natural and agro eco-systems in India. *Science International,* **1**, 124-131.

Kumar PS, Rabindra RJ and Ellison CA (2008). Expanding classical biological control of weeds with pathogens in India: the way forward. In: Proceedings of the XII International Symposiyum on Biological Control of Weeds, La Grande Motte, France

Rao, R.S., 1956. Parthenium – a new record for India. *Journal of the Bombay Natural History Society,* **54**, 218-220.

Simberloff D., Martin J.L., Genovesi P., Maris V., Wardle D.A., Aronson J., Courchamp F., Galil B., García Berthou E., Pascal M., Pyšek P., Sousa R., Tabacchi E. & Vilà M. (2013). Impacts of biological invasions: what's what and the way forward. *Trends in Ecology and Evolution* 28(1):58–66.http://dx.doi.org/10.1016/j.tree.2012.07.013

Sushil Kumar. (2014). Spread, menace and management of Parthenium. *Indian Journal of Weed Science,* **46(3)**, 205-219.

Vartak VD (1968). Weed that threatens crops and grasslands in Maharashtra. Indian Farming, **18**, 23-24.

Witten I.H. and Frank E.(2005). Data Mining: Practical Machine Learning Tools and Techniques, second ed. Elsevier, San Francisco.