



Report On:

“Machine learning based potential distribution of Siam weed in Bangladesh”

The report has been submitted to the Faculty of **Applied Statistics and Data Science**, **Jahangirnagar University** in partial fulfillment of the requirements for the Award of Masters of a professional studies (MPS) in **Applied Statistics and Data Science**, **Jahangirnagar University**.

Supervised By:

Prof. Md. Abdus Salam

Professor

Department of Applied Statistics and Data Science

Jahangirnagar University (JU)

Prepared By:

Md. Yamin Hossain

ID NO: 20228066

Batch: 8th (PM-ASDS)

Applied Statistics and Data Science (PM-ASDS)

Jahangirnagar University (JU)

LETTER OF TRANSMITTAL

Date: 22.09.2023

Prof. Md. Abdus Salam

Professor

Department of Applied Statistics and Data Science (PM-ASDS), Jahangirnagar University

Subject: Machine learning based potential distribution of Siam weed in Bangladesh

Dear Sir,

With due respect and humble submission, I would like to thank you for giving me a chance to prepare a report on this topic, “Machine learning based potential distribution of Siam weed in Bangladesh”. I have prepared the report highlighting the different schemes and analyze it from a research perspective. I have tried my best to make the report excellent and gather relevant information for preparing a complete report. Without the proper guidance of you, it was not possible for me to prepare the report.

I therefore, hope that this report will provide necessary information about “Machine learning based potential distribution of Siam weed in Bangladesh”. I will be available at any time convenient for further clarification on the report.

Sincerely yours,

Md. Yamin Hossain

ID NO: 20228066

Batch: 8th (PM-ASDS)

Applied Statistics and Data Science (PM-ASDS)

LETTER OF AUTHORIZATION

This is to certify that Md. Yamin Hossain, a student of PM-ASDS Program, ID No: 20228066 has successfully completed his thesis project entitled “Machine learning based potential distribution of Siam weed in Bangladesh” under my supervision as the partial fulfillment for the award of professional studies (MPS). He has done his job according to my supervision and guidance. He has tried his best to do this successfully. I think this program will help him in the future to build up his career. I wish him success and prosperity.

.....

Prof. Md. Abdus Salam

Professor

Department of Applied Statistics and Data Science

Jahangirnagar University (JU)

Table of Content

Acknowledgement	1
1 Introduction	2
2 Objective	4
3 Materials and Methods	5
3.1 Siam weed occurrence data	5
3.2 WorldClim bioclimatic data	5
3.3 Selection of predictors for model building	7
3.4 Model building framework	8
4 Results	10
4.1 Model accuracy	10
4.2 Maxent Model Validation	11
4.3 Model tuning with cross-validation	12
4.4 Variable importance	13
4.5 Potential distribution of Siam weed	14
5 Discussion	15
5.1 Limitation of Future Work	16
6 Conclusions	17
References	18
Source Code	19

List of Figures

Figure 1. Flowers and leaf of <i>Chromolaena odorata</i> (Siam weed).....	4
Figure 2. Black dots showing occurrence points	6
Figure 3. The 19 bioclimatic variables as image files collected	8
Figure 4. PlotROC RF.....	11
Figure 5. PlotROC Maxent.....	11
Figure 6. AUC and ROC for the different models.....	12
Figure 7. Flowchart showing methodology used for the predictive modeling of R.....	13
Figure 8. Few Infestations Area (RF).....	15
Figure 9. High Infestations Area (Maxent).....	15

List of Tables

Table 1. Bioclimatic and topographic variable names and description of the variables.....	7
Table 2. Table showing variance inflation factor (VIF)	14

Acknowledgement

I would like to express my gratitude to all the faculty and staff of the Department of Applied Statistics and Data Science at the Jahangirnagar University for their support during my master's degree. I thank Professor Dr. Abdus Salam for his guidance and input throughout this work.

I also recognize the Global Biodiversity Information Facility (GBIF) and the WorldClim project for providing valuable data for this research.

Finally, I also express my gratitude to my parents and friends for supporting me on this journey and motivating me throughout the research work.

.....

Author

Md. Yamin Hossain

Machine learning based potential distribution of Siam weed in Bangladesh

Md. Yamin Hossain

Abstract

Chromolaena odorata, commonly known as Siam weed, is an invasive plant species that poses a severe threat to biodiversity and agricultural productivity. Despite potential Siam weed infestation risks at the south-east and south-west regions of Bangladesh, no studies exist that document Siam infestation. Using Random Forest and maximum entropy (Maxent) modeling approach, the present study tries to address the problem in the concerned region comprising three divisions: Chittagong, Khulna and Rajshahi. The invasive nature of this weed poses a significant threat to the biodiversity and agricultural productivity of the country. To mitigate these impacts, understanding its distribution patterns is crucial. This research aims to predict the current potential distribution of *Chromolaena odorata* (Siam weed) in Bangladesh using machine learning modeling techniques, specifically Random Forest and Maxent modeling. Through the analysis of various environmental and geographic factors, this study seeks to generate accurate predictive models. The results will aid in developing effective management strategies and policies for controlling the spread of *Chromolaena odorata* in Bangladesh.

1 Introduction

Chromolaena odorata, commonly known as Siam weed, is an invasive plant species that poses a severe threat to biodiversity and agricultural productivity. In recent years, the invasive plant species has become a significant concern for agriculture, biodiversity, and ecosystems in Bangladesh. Siam weed belongs to the Asteraceae family, is a native plant of North America, Central America and South America, is one of the world's worst weeds. It is a highly invasive weed with a wide range of negative impacts on both the environment and agriculture. It spreads rapidly and can quickly dominate native vegetation, outcompeting other plants for resources such as sunlight, water, and nutrients. It also poses health risks to humans and animals, as it can cause skin irritations and allergies upon contact. Siam weed can spread rapidly and can lead to the displacement of native flora and fauna, resulting in ecological imbalances and economic losses for the agricultural sector. To combat this threat, it is essential to understand the current and potential distribution of *Chromolaena odorata* within the country.



leaf



Flowers

Figure 1. Flowers and leaf of *Chromolaena odorata* (Siam weed). (Source: <https://www.cabidigitallibrary.org>)

However, to conduct such a through survey throughout the country is extremely time consuming, labor intensive and can be very costly. In the case of small areas, conducting through field surveys can be enough to understand the nature and magnitude of the infestation of invasive weed species. At a large scale, such as the whole country level, it may not be feasible. That is why species distribution modeling (SDM) can be very useful. Application of SDM often involves collection of weed occurrences by conducting field surveys.

Where, occurrences of a species are recorded as *presence* and the longitude and latitude of that location is recorded. On the other hand, if the species is absent on a location, the location information (also longitude and latitude) are recorded as *absences*. This *absence* and *presence* points of an entire study area can be dummy coded as zeros (0) and ones (1). Then to understand what actually caused such a distribution of a species, we can use several predictors for example, temperature, rainfall, elevation of these locations to make models. So, in simple terms to make a SDM we need to collect presence and absence points, using a handheld GPS device, and also need to collect predictors (rainfall, temperature, elevation etc).

If we have both presence and absence points, we can make generalized linear models where our *presence/absence* points in one column are dependent variables (Y). And the predictors, for examples temperature (X_1), rainfall (X_2) to n as predictors.

$$Y \sim X_1 + X_2 + \dots + X_n$$

However, we only know the presence of a species and we lack the absence points. Therefore, it is very difficult to use a generalized linear model (GLM) framework. Hence, machine learning models such as random forest (RF), maximum entropy modeling (Maxent) were developed and are very popular in building SDM. Particularly, the Maxent algorithm revolutionized the ecological modeling domain. The main reason why the Maxent algorithm is very popular, is that if we only the presence locations, and roughly assume the absences, we can still make a good prediction about a species distribution. Recently RF model has also shown a good performance in prediction species distribution, however it needed to be tested dependent sites and other consideration,

Very few studies have been done of the invasive weed species distribution in Bangladesh. Masum et al., (2022) used Maxent algorithm to predict the potential distribution of Parthenium, an invasive weed species, in Bangladesh. Although several regional or global studies cover Bangladesh in their studies, such regional models can be unreliable without considering the critical details. Furthermore, Siam weed is also a common pest in agricultural lands of Bangladesh, and there are no studies showing how this species distribution impacts the agricultural ecosystems. Furthermore, without such models, it is also impossible to understand the future distribution of Siam weed under climate change scenarios. To fill these research gaps, this research project tests the two popular machine learning algorithms: (i) Random Forest (RF) and, (ii) Maxent

These two predict Siam weeds potential distribution. It can be a valuable tool for policymakers and environmentalists for creating targeted management strategies to control the spread of *Chromolaena odorata* and mitigate its impact on Bangladesh's ecosystems and agriculture.

2 Objective

The main aim of this study is to develop a potential species distribution map of Siam weed in Bangladesh. To achieve that I use the data Siam weed occurrence data from the Global Biodiversity Information Facility (GBIF) website, along with environmental predictors from the WorldClim website. The specific objectives are:

- (1) To test two machine learning algorithms, random forest and Maxent, for predicting Siam weed species distribution in Bangladesh.
- (2) To find out the potential vulnerable areas of Siam weed species in Bangladesh

3 Materials and Methods

3.1 Siam weed occurrence data

The occurrence data of Siam weed were collected from the Global Biodiversity Information Facility (GBIF) website (www.gbif.org). GBIF is a global hub for biodiversity data, enabling easy access and sharing of information about species occurrences and distributions. So, I collected species occurrence data for *Chromolaena odorata* from the GBIF website (Lane and Edwards, 2017) using the *spocc* package library (Owens et al., 2023) in R open-source software version 4.3.1 (R Core Team, 2023). This data is invaluable for my research as it provides insights into the global distribution of Siam weed.

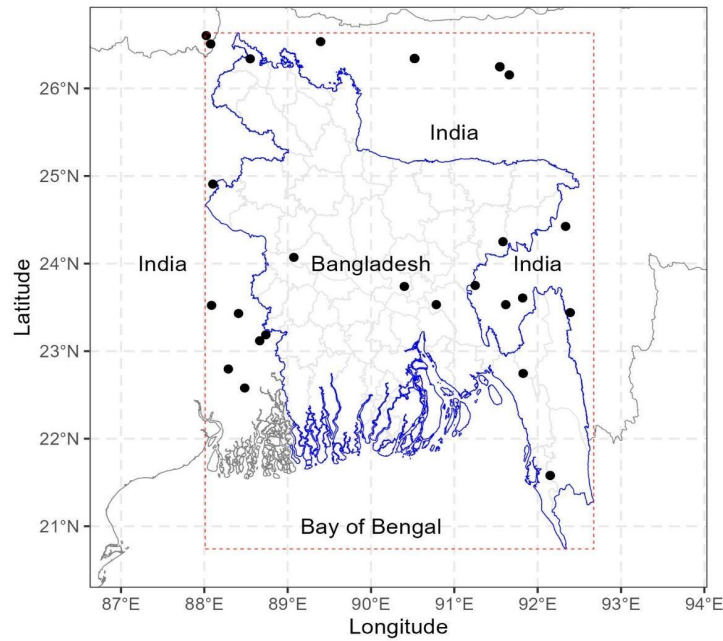


Figure 2. Black dots showing occurrence points (n=40) of Siam weed collected from GBIF database (www.gbif.org). The red dotted line is the extent of this study area.

3.2 *WorldClim bioclimatic data*

To understand the distribution of Siam weed, I used the 19 bioclimatic data from the WorldClim project (Fick and Hijmans et al., 2017; available at www.worldclim.org). First, I downloaded all 19 variables for Bangladesh that were available as raster files using the *geodata* package (Hijmans et al., 2023) in R software.

These data were generated from various meteorological stations around the world for the year 1970 to 2000. This represents the current environmental conditions. Table 1 shows the description of the 19 bioclimatic variables which were used as predictors for modeling the distribution of Siam weed. Fig. 2 shows the maps of the 19 bioclimatic variables plotted for the study area.

Table 1. Bioclimatic and topographic variable names and description of the variables initially chosen for modeling.

SL	Code	Name	Unit	Source
1	Bio1	Annual mean temperature Mean diurnal range	°C	WorldClim
2	Bio2	(Mean of monthly (max temp–in temp))	°C	
3	Bio3	Isothermality (Bio2/Bio7) (× 100)	°C	
4	Bio4	Temperature seasonality (Standard deviation × 100)	°C	
5	Bio5	Max temperature of warmest month	°C	
6	Bio6	Min temperature of coldest month	°C	
7	Bio7	Temperature annual range (Bio5–Bio6)	°C	
8	Bio8	Mean temperature of wettest quarter	°C	
9	Bio9	Mean temperature of driest quarter	°C	
10	Bio10	Mean temperature of warmest quarter	°C	
11	Bio11	Mean temperature of coldest quarter	°C	
12	Bio12	Annual precipitation	mm	
13	Bio13	Precipitation of wettest month	mm	
14	Bio14	Precipitation of driest month	mm	
15	Bio15	Precipitation seasonality (Coefficient of variation)	1	
16	Bio16	Precipitation of wettest quarter	mm	
17	Bio17	Precipitation of driest quarter	mm	
18	Bio18	Precipitation of warmest quarter	mm	
19	Bio19	Precipitation of coldest quarter	mm	

3.3 Selection of predictors for model building

Using all the 19 variables a SDM model can be made however as these variables can be highly correlated to each other. In such a case, it is better to omit the correlated variables. To address the potential issue of collinearity, a threshold for Pearson correlation coefficient was set at 0.70. To select predictors with low correlation I used the R package library *usdm* (Naimi et al., 2015). This means that variables that had higher correlation coefficient than 0.70, as set by the Variable Inflation Factor (VIF) threshold in the *usdm* package were considered multicollinear. Such variables were then omitted from the model building process. A total of 6 variables were ultimately chosen for modeling.

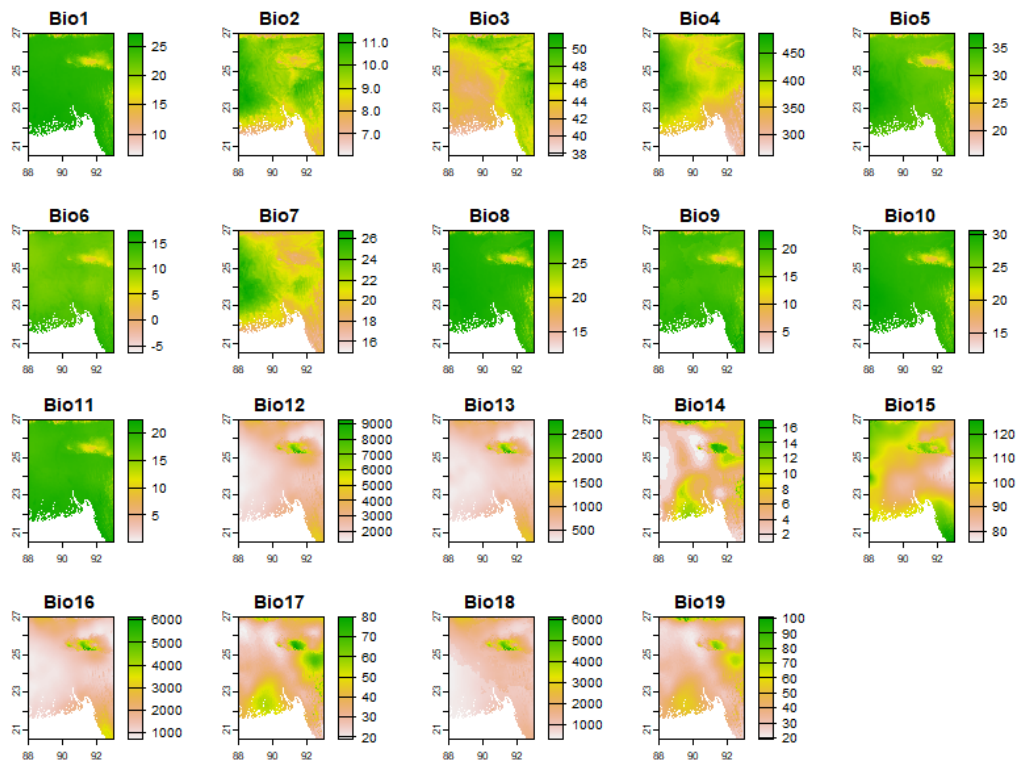


Figure 3. The 19 bioclimatic variables as image files collected from the WorldClim website (www.worldclim.org).

The remaining size predictor variables were of the class *SpatRaster* class in the R package library *terra* (Hijmans et al., 2022), which represents a three-dimensional grid with dimensions of 780 rows, 600 columns, and 6 layers. Each layer corresponds to a distinct bioclimatic variable, viz Bio2, Bio3, Bio9, Bio15, Bio18, Bio19.

3.4 Model building framework

For building a machine learning based species distribution model, I used the SDMtune library (Vignali et al., 2020) in the R open-source software. SDMtune provides a user-friendly framework that enables the training and the evaluation of species distribution models (SDMs). Using this SDMtune library I can seamlessly fit four different machine learning models viz. i) artificial neural network (ANN), (ii) bootested random tree (BRT), (iii) maximum entropy model (Maxent) and, (iv) random forest (RF).

For this study, I have selected two machine learning algorithms i) random forest and ii) Maxent. The Maxent algorithm is one of the popular algorithms used for SDMs (Phillips, 2005; Elith et al., 2011; Masum et al., 2022). On the other hand, random forest is perhaps the most used algorithm nowadays, both for regression and classification purposes. However, recent study also shows that this can be well adapted in building a machine learning based SDM (Valavi et al., 2021).

a) Random Forest Model (RF)

I collected 1000 random absence points using the `spatSample()` function within the Bangladesh area. During this command, I set a seed to reproduce the randomly generated point coordinates. Subsequently, I prepared a Species With Data (SWD) object for model building by employing the `prepareSWD()` function. This SWD object was created with the species name *Chromolaena odorata*, a prediction data frame containing the collected presence locations, and a data frame with randomly generated absence locations. Previously selected uncorrelated six environmental variables were used as predictors for the model.

I divided the presence locations into training (80%) and testing (20%) datasets for model evaluation. There were 40 presence locations and 1,000 absence locations in the dataset. Then using the `train()` function a default model was built. I ran this SDMtune `train()` function with its default attributes, which included 500 trees in the forest and a random selection of `mtry` (the square root of the number of predictors) variables at each split.

However, the obtained results were unreliable. The training AUC was nearly hundred percentage, which implies that the model fit the training data almost perfectly. In reality, such a high training AUC is improbable, indicating potential overfitting to the dataset. Subsequently, I attempted k-fold cross-validation and hyperparameter tuning with default parameters, but the training and test AUC values did not change significantly. Furthermore, my predictive map did not provide sufficient results.

b) Maxent Modeling

After the previous Random Forest model did not perform satisfactorily, I made the decision to transition to a different machine learning approach. I opted to use Maxent, a relational machine learning model commonly employed for predicting cropland suitability. Once again, I began by preparing a SDM object, this time with the intention of constructing a Maxent model within the R environment.

For model evaluation purposes, I divided the presence locations into training (80%) and testing (20%) datasets. There were 40 presence locations and 1,000 absence locations in the dataset. To create the Maxent model, I employed the `SDMtune train()` function, which is implemented using the Maxent package (Phillips 2017). By employing the Maxent model and adjusting its attributes as needed, I aimed to improve the model's performance and suitability for predicting cropland suitability.

Below the settings used to train the Maxent model:

- Train locations: 1032
 - presence: 32
 - absence/background: 1000
- Test locations: 1008
 - presence: 8
 - absence/background: 1000
- Continuous variables: Bio2, Bio3, Bio9, Bio15, Bio18, Bio19
- Output type: cloglog
- Feature Class combination: lqpht
- Regularization multiplier: 0.3
- Do clamping for predictions: TRUE

All the data processing and model building were performed using the open-source software R version 4.3.1 (R Core Team, 2022).

4 Results

4.1 Model accuracy

Among the two models, the RF model showed unreliable predictions. Although the RF model initially appeared to outperform the Maxent model, as demonstrated by its good test AUC value (AUC = 0.943), the Maxent model ultimately proved to be more satisfactory when predicting the final map. Despite efforts to improve the random forests model and prediction map, its performance remained subpar. Consequently, the Maxent model was selected for further processing..

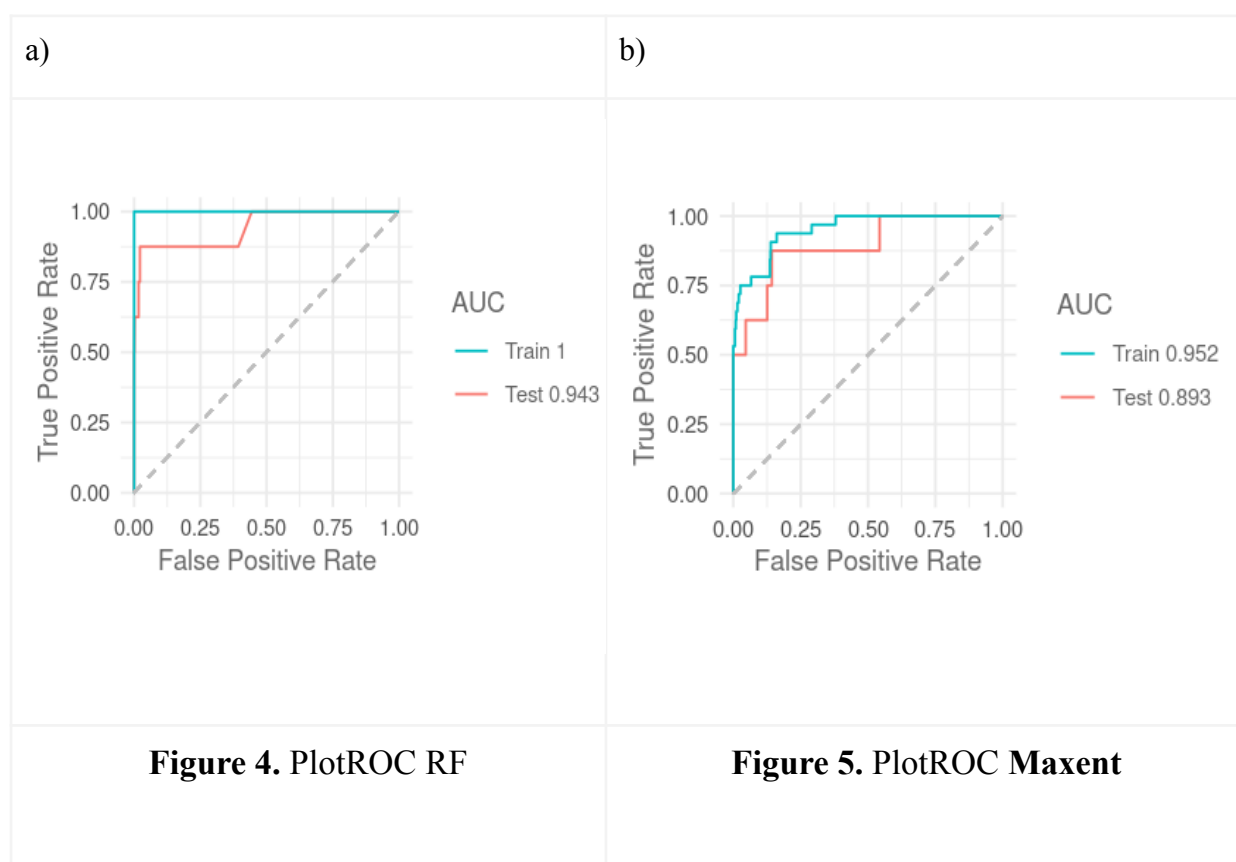


Figure 4. the performance of the Random Forest (RF) model by plotting its ROC curve and quantifying its performance using the Area Under the Curve (AUC) metric. The AUC for this model reflects its ability to distinguish between positive and negative classes, with higher values indicating better discrimination.

In Figure 5. We assess the performance of the Maxent model using the ROC curve and AUC. This metric quantifies the model's ability to make accurate binary predictions. The AUC value for Maxent provides valuable insights into its classification accuracy and separation of classes.

4.2 Maxent Model Validation

I run a total of 150 models, each with various combinations of regularization multipliers (RM) and feature class (FC) settings. Figure 6 illustrates the range of AUC values obtained across these models, showcasing the impact of different RM values on model performance. After extensive optimization efforts, only one model met my selection criteria, achieving the highest test AUC value. This particular model had a feature class combination of LGPHT and a specific regularization multiplier (RM) value, demonstrating its superiority in the context of my study.

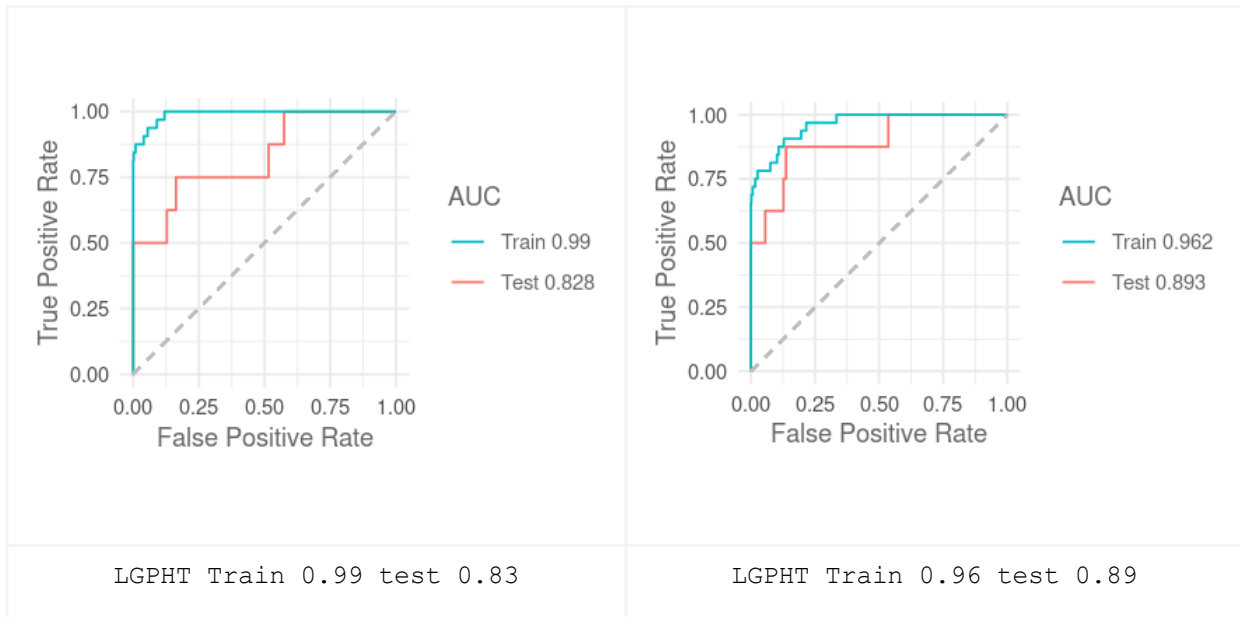


Figure 6. AUC and ROC for the different models

An AUC (Area Under the Curve) value of 0.89 in the ROC (Receiver Operating Characteristic) analysis is often considered indicative of a reasonably good model fit. This suggests that the model's ability to discriminate between the positive and negative classes is fairly effective, with an 89% chance of the model correctly ranking a randomly chosen positive instance higher than a randomly chosen negative instance.

4.3 Model tuning with cross-validation

How do I evaluate models incorporating a series of combinations of regularization multiplier (RM) values? A simple flowchart given below.

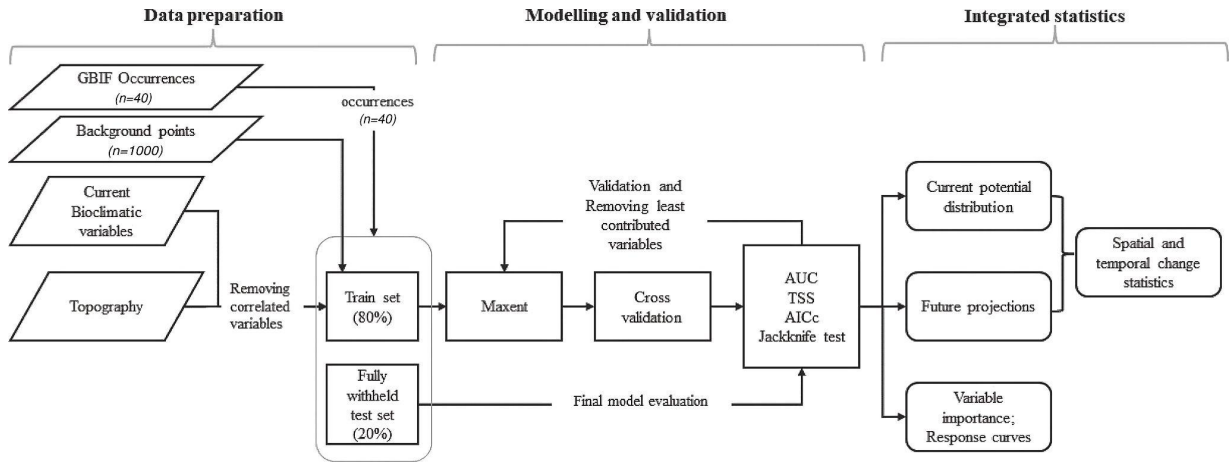


Figure 7. Flowchart showing methodology used for the predictive modeling of R. Siam using Maxent modeling. (Source: Mandal et al., 2023)

I conducted an extensive grid search to fine-tune the model parameters, utilizing the mean area under the receiver operating characteristic (ROC) curve (AUC) as the primary performance metric. In total, I explored 150 cross-validated model combinations by varying three regularization multiplier (RM) values and five feature classes (FC).

The five feature classes, namely L (linear), Q (quadratic), H (hinge), P (product), and T (threshold), represented a wide spectrum of feature engineering approaches. This comprehensive exploration allowed me to assess the impact of different RM values and feature classes on model performance.

Through this meticulous grid search, the goal was to pinpoint the optimal model configuration that would yield the highest test AUC, indicative of the model's ability to discriminate between positive and negative instances. This systematic approach ensures that the selected model is well-suited for the specific task and dataset, delivering robust and reliable results for further analysis or application.

4.4 Variable importance

Among the predictor variables, Mean of monthly max temp–in temp (Bio2); Precipitation of coldest quarter (Bio19); Mean temperature of driest quarter (Bio9); Isothermality (Bio2/Bio7) ($\times 100$) (Bio3); Precipitation of warmest quarter (Bio18); and Precipitation seasonality (Coefficient of variation)(Bio15) accounted for more than a 90% contribution to the model (Table 2).

Table 2. Table showing variance inflation factor (VIF)

Variable	Permutation_importance	sd
Bio2	31.6	0.022
Bio19	21.3	0.027
Bio9	13.7	0.015
Bio3	13.0	0.011
Bio18	11.3	0.019
Bio15	9.0	0.015

Among the six predictors used for Maxent modeling, the Mean of monthly max temp–in temp (Bio2) accounted for approximately 31 % of the permutation importance [Table 2], and similarly mean precipitation of coldest quarter (Bio19) accounted for approximately 21 % [Table 2]. This indicated that Bio2 and Bio19 are very important in predicting habitat suitability for the species.

This finding was reflected in jackknife tests for variable importance, which showed that Bio19 was the most important variable for model information gain (Figure 9). Other important predictor variables were Precipitation of the warmest quarter (Bio18) and mean Precipitation seasonality Coefficient of variation (Bio15).

4.5 Potential distribution of Siam weed

Based on the Maxent model, it indicates that three divisions in the south-east and south-west parts of Bangladesh were highly infested with siam weed. The first division showing a high potential for infestation is Chittagong, including the districts of Chattogram, Rangamati, Bandarban, and Khagrachari, where probabilities reached 1.00. These areas are considered to have the highest infestations and are highly suitable for siam weed.

The Maxent model also predicts lower probabilities of infestations in the southern and western parts of the map, with probabilities reaching 0.00 in some areas. This suggests that these regions are at the lowest risk of siam weed infestations.

Final Results RF and Maxent Modeling Approach

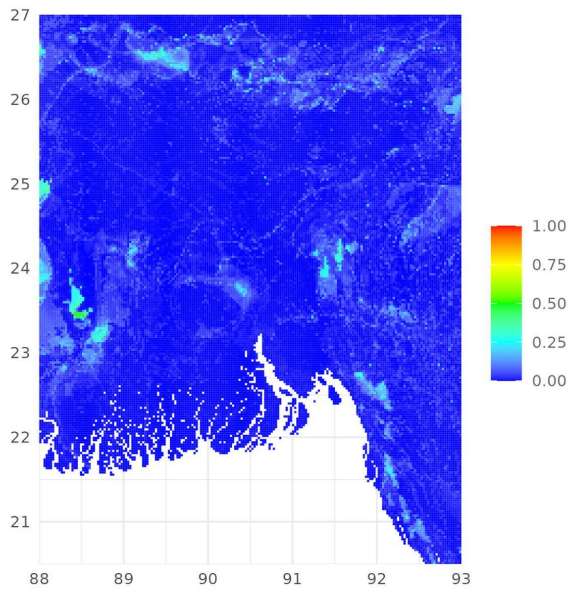


Figure 8. Few Infestations Area (RF)

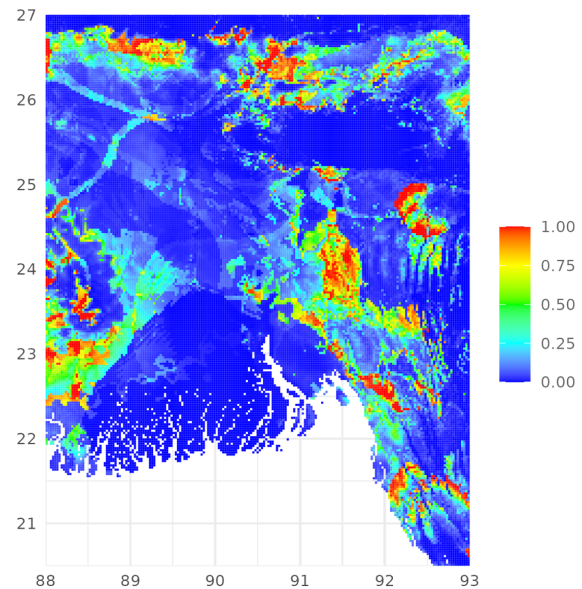


Figure 9. High Infestations Area (Maxent)

Under the current climatic conditions and based on occurrence records, potential habitat suitability maps are depicted in Figure 9. According to the Maxent model, the study region includes Chittagong, Rangamati, Khagrachari, Feni, Comilla, Satkhira, Khulna, Magura, Narail, and Jessore Districts. The southeastern and eastern hill upazilas of Chittagong demonstrate the highest suitability for siam weed, while the northwestern districts Magura, Narail, and Jessore show the lowest suitability for the weed.

The variation in habitat suitability across different districts can be attributed to a range of environmental factors. It's worth considering how climate change and other environmental factors might influence siam weed distribution in the future. Climate models and projections can help anticipate potential shifts in suitability and guide adaptive management strategies.

5 Discussion

Siam weed (*Chromolaena odorata*) is a major invasive plant species causing significant concerns in several countries, particularly across parts of Asia. This aggressive plant's rapid and extensive spread is attributed to its adaptability to various environments and prolific reproduction. Siam weed tends to flourish in disturbed areas such as agricultural lands, forest clearings, and roadsides, quickly outcompeting native vegetation. Its invasive nature has adverse effects on biodiversity, altering habitat structures and reducing plant diversity.

The findings presented in this study provide crucial insights into the current habitat suitability of siam weed in Bangladesh, particularly in the southeastern and southwestern regions. The Chittagong division, encompassing districts such as Chattogram, Rangamati, Bandarban, and Khagrachari, emerges as the most highly infested area, with probabilities reaching a significant 1.00. This indicates a high risk of siam weed infestation in these regions.

Conversely, the southern and western parts of the map exhibit lower probabilities of infestations, with some areas even reaching a probability of 0.00. This suggests that these regions face a considerably lower risk of siam weed infestations. These findings are consistent with the potential habitat suitability maps generated through Maxent modeling approaches, further reinforcing the robustness of the results.

The identification of Chittagong division as a high-risk area aligns with previous studies highlighting the region's susceptibility to invasive plant species. The presence of siam weed in this area is likely influenced by a combination of environmental factors specific to these districts, which may include climate, soil type, land use patterns, and other ecological variables.

In Figure 8, we have maps showing where siam weed can thrive based on the current climate and where it's found. According to the RF (Random Forest) model, this includes places like Chittagong, Rangamati, Khagrachari, Feni, Comilla, Satkhira, Khulna, Magura, Narail, and Jessore Districts.

In simple terms, siam weed seems to like the southeastern and eastern parts of Chittagong a lot, but it's not so fond of the northwest areas like Magura, Narail, and Jessore. Why these differences? It's because of different things in the environment.

Now, let's think about the future. How will climate change and other things in nature affect where siam weed can grow? To figure this out, we can use special computer models that help us predict the future. These predictions can help us make plans and decisions to manage siam weed better

The study also highlights the variability in habitat suitability across different districts. Factors contributing to this variation could range from local climate patterns to specific land management practices. It is essential to delve deeper into these localized factors to better understand and mitigate siam weed infestations.

Looking ahead, it is imperative to consider the potential impacts of climate change and other environmental shifts on siam weed distribution. By incorporating climate models and projections into future assessments, we can anticipate potential shifts in suitability and proactively implement adaptive management strategies. This forward-looking approach will be vital in effectively managing and controlling siam weed infestations in Bangladesh. This study provides a comprehensive assessment of siam weed habitat suitability in Bangladesh, shedding light on high-risk areas such as the Chittagong division, and contrasting them with regions at lower risk.

The insights gleaned from this research serve as a valuable foundation for informed decision-making and the development of targeted management strategies to combat siam weed infestations in the country.

5.1 Limitation of Future Work

While this study provides valuable insights, it has certain limitations. The accuracy of the models depends on the quality of the input data from GBIF. Therefore, future studies should focus on conducting field surveys and combining the data available from GBIF.

Future work related to siam weed (*Chromolaena odorata*) is the need for more comprehensive research on its genetic diversity and potential for evolving resistance to control measures. Understanding the genetic variability within siam weed populations is crucial for developing effective management strategies.

Additionally, studying the weed's capacity to adapt to changing environmental conditions, especially in the context of climate change, is essential. Long-term monitoring and modeling studies are necessary to anticipate and address potential shifts in siam weed distribution and behavior. Furthermore, research should explore sustainable and eco-friendly control methods to minimize the environmental impact of eradication efforts.

6 Conclusions

In conclusion, this study has provided valuable insights into the habitat suitability and potential infestation risk of siam weed in various regions of Bangladesh. The key findings and conclusions drawn from this research are as follows:

1. **High-Risk Areas:** The Chittagong division, encompassing districts like Chattogram, Rangamati, Bandarban, and Khagrachari, has been identified as a high-risk area for siam weed infestations. The probabilities of infestation in these regions reached the maximum value of 1.00, indicating a significant threat.
2. **Low-Risk Areas:** Conversely, the southern and western parts of the study area exhibited lower probabilities of siam weed infestations, with some areas registering probabilities as low as 0.00. These regions are considered to be at the lowest risk of siam weed infestations.
3. **Model Validation:** The consistency of results obtained through both the Random Forest (RF) and Maxent modeling approaches adds credibility to the findings, demonstrating the reliability of the predictive models employed in this study.
4. **Environmental Factors:** The variation in habitat suitability across different districts can be attributed to a range of environmental factors. These factors may include climate, soil characteristics, land use patterns, and other ecological variables. Further research is needed to comprehensively understand these localized factors influencing siam weed distribution.
5. **Future Considerations:** As climate change and other environmental factors continue to evolve, it is crucial to consider how these changes might impact siam weed distribution in the future. Integrating climate models and projections into future assessments will enable us to anticipate potential shifts in suitability and guide adaptive management strategies.

In summary, this research contributes significantly to our understanding of siam weed distribution and infestation risk in Bangladesh. It not only identifies areas of concern but also highlights the importance of ongoing monitoring and adaptive management to address the challenges posed by invasive plant species. Implementing proactive measures based on these findings will be essential to safeguard the agricultural and ecological landscapes of Bangladesh in the face of changing environmental dynamics.

References

1. Fick, S. E., & Hijmans, R. J. (2017). WorldClim 2: new 1-km spatial resolution climate surfaces for global land areas. *International journal of climatology*, 37(12), 4302-4315.
2. Hijmans, R.J., Barbosa, M., Ghosh, A., Mandel, A. (2023). geodata: Download Geographic Data. R package version 0.5-8, <https://CRAN.R-project.org/package=geodata>
3. Lane, M. A., & Edwards, J. L. (2007). The global biodiversity information facility (GBIF). *Systematics Association special volume*, 73, 1.
4. Owens, H., Barve, V., & Chamberlain, S. (2023). spocc: Interface to species occurrence data sources. *R package version 1.2.2*. See <http://CRAN.R-project.org/package=spocc>.
5. Naimi, B. (2015). USDM: Uncertainty analysis for species distribution models. R package version 1.1–15. R Documentation.
6. Vignali, S., Barras, A. G., Arlettaz, R., & Braunisch, V. (2020). SDM tune: An R package to tune and evaluate species distribution models. *Ecology and Evolution*, 10(20), 11488-11506. <https://doi.org/10.1002/ece3.6786>
7. Phillips, S. J. (2005). A brief tutorial on Maxent. *AT&T Research*, 190(4), 231-259.
8. Elith, J., Phillips, S. J., Hastie, T., Dudík, M., Chee, Y. E., & Yates, C. J. (2011). A statistical explanation of MaxEnt for ecologists. *Diversity and distributions*, 17(1), 43-57.
9. Valavi, R., Elith, J., Lahoz-Monfort, J. J., & Guillera-Aroita, G. (2021). Modeling species presence-only data with random forests. *Ecography*, 44(12), 1731-1742.
10. Hijmans, R. J., Bivand, R., Forner, K., Ooms, J., Pebesma, E., & Sumner, M. D. (2022). Package ‘terra’. Maintainer: Vienna, Austria.
11. R Core Team. (2023). R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria. <https://www.R-project.org/>.
12. Vignali, S., Barras, A. G., Braunisch, V. (2023). SDM tune: Species Distribution Model Selection. R package version 1.3.1. <https://CRAN.R-project.org/package=SDMtune>

Source Code

For view source code or download online pdf version, please scan below QR-Code



Above QR For Project Source Code

~ YAMiN ~

For Online PDF Version, Scan This QR Code



Above QR For Download PDF Book