Assessment of American Bullfrog (Lithobates catesbeianus) spreading in the Republic of Korea using rule learning of elementary cellular automata

Gyujin Oh
Chonnam National University
Yunju Wi
Chonnam National University
Hee-Jin Kang
Chonnam National University
Seung-ju Cheon
Chonnam National University
Ha-Cheol Sung
Chonnam National University
Yena Kim
Hawaii Pacific University
Hong Sung Jin (hjin@jnu.ac.kr)
Chonnam National University

Article

Keywords: Bullfrogs Spreading, Clustering, ECA, CNN, BOS, Habitat Suitability

Posted Date: November 8th, 2023

DOI: https://doi.org/10.21203/rs.3.rs-3545672/v1

Additional Declarations: No competing interests reported.

Assessment of American Bullfrog (Lithobates catesbeianus) spreading in the Republic of Korea using rule learning of elementary cellular automata

Abstract

The spread of American Bullfrog, one of the 100 of the World's Worst Invasive Alien Species, has a great impact on the surrounding ecosystem. Little is known about the tendancy and pattern of how they are spreading in South Korea geographically. It is important to study the tendancy of their spreading so that a proper mitigation can be applied when needed. This study is based on the results of national surveys that observed the distribution. The entire data is divided into 25 regional clusters using the divisive hierarchical clustering method. In order to estimate the degree of spreading, a sequence of spatial distribution is constructed for each cluster using the agglomerative clustering method. ECA(elementary cellular automata) is introduced to find rules governing the pattern variation in the sequence. Each cell represents either the observed or unobserved site of bullfrog. The number of Bullfrog Observed Site (BOS) in a sequence of each cluster is counted and used to define the spreading intensity. The rules of ECA are learned by the CNN(Convolution Neural Network) method and used to estimate and predict the spreading intensity by counting the expected number of BOS over 400 generations. Taking environmental factors into account, habitat suitability is used and obtained using Maxent. The spreading intensity is multiplied by the habitat suitability to get an assessment of bullfrogs spreading. The relative spreading assessment is estimated, which is classified into 4 groups; spreading intensively, spreading slowly, maintaining or declining population.

Keywords : Bullfrogs Spreading, Clustering, ECA, CNN, BOS, Habitat Suitability

Gyujin Oh[1], Yunju Wi[1], Hee-Jin Kang[2], Seung-ju Cheon[2], Ha-Cheol Sung[3], Yena Kim[4], Hong-Sung Jin[1*]

1 Department of Mathematics & Statistics, Chonnam National University, 77 Yongbongro, Bukgu, Gwangju 61186, Republic of Korea
2 School of Biological of Sciences and Biotechnology, Chonnam National University, 77 Yongbongro, Bukgu, Gwangju 61186, Republic of Korea
3 Department of Biological Sciences, College of Natural Sciences, Chonnam National University, 77 Yongbongro, Bukgu, Gwangju 61186, Republic of Korea
4 Department of Mathematics, Hawaii Pacific University, 1 Aloha Tower Drive, Honolulu, HI 96813
* Corresponding author:
E-mail: hjin@jnu.ac.kr
Phone: +82 62 530 3346

1 Introduction

The American bullfrog, *Lithobates catesbeianus*, has been introduced to more than 40 countries worldwide and is listed on the "100 of the World's Worst Invasive Alien Species" (Database 2023). American Bullfrog was introduced to Korea in 1957 and cultivated for the purpose of establishing new food sources for human consumption, but due to its low economic efficiency and low demand as food, most farm gave up on farming and released them into rivers illegally, and bullfrogs were spread all over the country (Jang and Suh 2010; Kim 2009; Oh and Hong 2007). The spread of bullfrogs has a great impact on the surrounding ecosystem such as increased competition with native species, predation and the spread of ranavirus (Ficetola et al. 2010; Ficetola et al. 2007; Giovanelli et al. 2008; Groffen et al. 2019; Iñiguez and Morejón 2012; Kamoroff et al. 2020; Koo and Choe 2021; Nori et al. 2011; Park et al. 2022; Schlaepfer et al. 2005). Bullfrogs have continued to spread in an environment without natural enemies and have now spread nationwide except in some inland (mountain) areas in South Korea (Kang et al. 2019; Koo and Choe 2021). The species was reported to occur at 2,716 sites, mainly along the southern and western coasts, but was rarely distributed in the northern part of Korea or along the eastern coast (Kang et al. 2019). It is also predicted that bullfrogs will continue to spread further in the future (Koo and Choe 2021). Although several management strategies were implemented, the effectiveness of the past control decisions is unclear (Chang et al. 2022). Meanwhile, some reports stated that local natural enemies have appeared and are controlling the bullfrog population (No et al. 2017).

In this study, the likelihood of future spread is assessed by calculating the intensity of spread and habitat suitability in 25 regions. Then they are classified into areas where the population is expected to continue to increase, areas where there is no significant change in the current population, and finally areas where the population is expected to decrease.

The study is based on the findings of national surveys including Natural Resources from 2006 to 2012, the National Wetland Center Report from 2011 to 2017 and the National Institute of Ecology from 2015 to 2017 (Kang et al. 2019).

Since we do not have time series data of bullfrog distribution, we analyze the spatial distribution using hierarchical divisive clustering method using scikit-learn 1.3.0 (da Silveira Vasconcelos et al. 2011; Ermentrout and Edelstein-Keshet 1993; Patlolla 2018; Pedregosa et al. 2011). The entire data is clustered into small clusters, and the degree of spreading is estimated by the evolution rules from the elementary cellular automata scheme (Nagatani and Tainaka 2018; Wolfram 1983, 2002) in each small cluster. CNN is trained to learn the evolution rules (Brodrick et al. 2019; Deneu et al. 2021; Duryea 2018; Kattenborn et al. 2021; Qin et al. 2020). By recognizing small clusters as a single image of 0's and 1's, the number of 1's is counted, which is the number of Bullfrog Observed Site (BOS). The ratio of the expected number of BOS at time *t* to the initial number of BOS is used to define the spreading intensity. The estimated spreading intensity is multiplied by the habitat suitability to express the assessment of bullfrog spreading by region. The habitat suitability is achieved using Maxent software (Elith* et al. 2006; Ficetola et al. 2010; Ficetola et al. 2007; Steven J. Phillips 2017; Steven J Phillips et al. 2017; Steven J Phillips et al. 2006; Steven J Phillips and Dudík 2008; Tesfamariam et al. 2022; Venne and Currie 2021).

2 Material and Methods

Following the procedure shown in Fig. 1, several machine techniques are used such as clustering, convolutional neural network, elementary cellular automata rule learning and Maxent to assess the spreading intensity of bullfrogs by region.

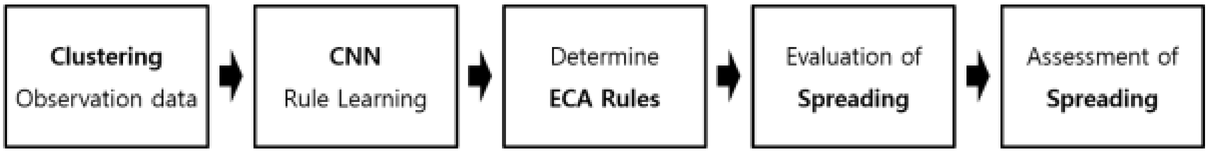


Figure 1. The process to get an assessment of spreading. For each cluster, the agglomerative clustering method is used to generate a clustering sequence, which is regarded as an image sequence, and the CNN method is applied to learn ECA rules governing the sequence variation. For each ECA rule the number of 1's is counted, which is the expected number of BOS to evaluate the intensity of spreading. Finally, the intensity is adjusted by multiplying habitat suitability to get an assessment of spreading

2.1 Observation Data

Data are collected from the results of several official nation-wide surveys, including the National Survey of Natural Resources from 2006 to 2012, the National Wetland Center Report from 2011 to 2017 and the National Institute of Ecology from 2015 to 2017 (Kang et al. 2019). Figure 2 shows the distribution of American Bullfrog observed in South Korea. Time series data is not given.

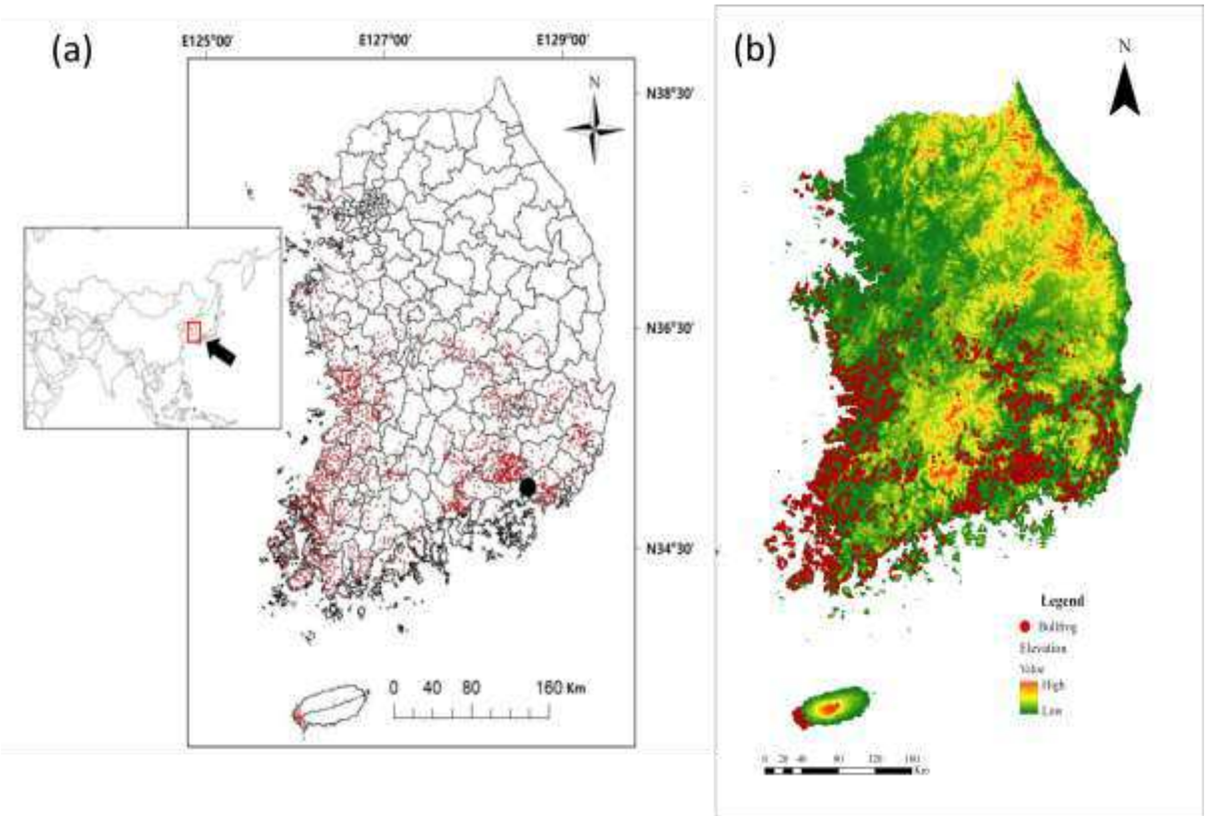


Figure 2. Distribution of bullfrog observations according to administrative districts and topography. The map above represents South Korea, and the data is between latitude 34°58’-36°71’and longitude 126°11’-128°2’, covering approximately the southern half of South Korea. **a** It shows where bullfrogs have been found on the maps with the boundaries of administrative districts. **b** It shows where bullfrogs have been found on the topographic map. The highest elevations are red, then moving to orange, yellow, bright greens and finally dull greens at the lower elevations. It is mainly distributed in coastal wetlands or riverside wetlands and is rarely distributed in mountainous areas. This is a collection of findings over 60 years, with no temporal information

2.2 Clustering

In order to estimate the intensity of spreading by region, a sequence of spatial distribution from the observed data in Fig. 2 is constructed using the divisive hierarchical clustering method. All observations start in one cluster of full data, and splits are performed recursively as one moves down the hierarchy by grouping neighboring data into the same cluster (Patlolla 2018). The scikit-learn clustering software (Pedregosa et al. 2011) is used, and clusters are numbered according to the order in which they are formed. Clustering is performed until 25 clusters are formed to roughly match the size of the administrative district. Rectangular images consisting of 20 by 20 cells are created by uniformly dividing the latitude and longitude including all observations in each cluster. Latitude and longitude information for all clusters is in Table 1. If each cell had a bullfrog observation point, it is marked as 1, otherwise it is marked as 0. Here, the point density of each cell is inhomogeneous. To estimate the spreading intensity of each cluster the agglomerate clustering method is performed in each cluster making the sequence of images, $C_0 \rightarrow C_1 \rightarrow \cdots C_{n-1} \rightarrow C_n$. Figure 3 illustrates the agglomerate clustering steps, taking cluster #5 in Fig. 6d as an example.

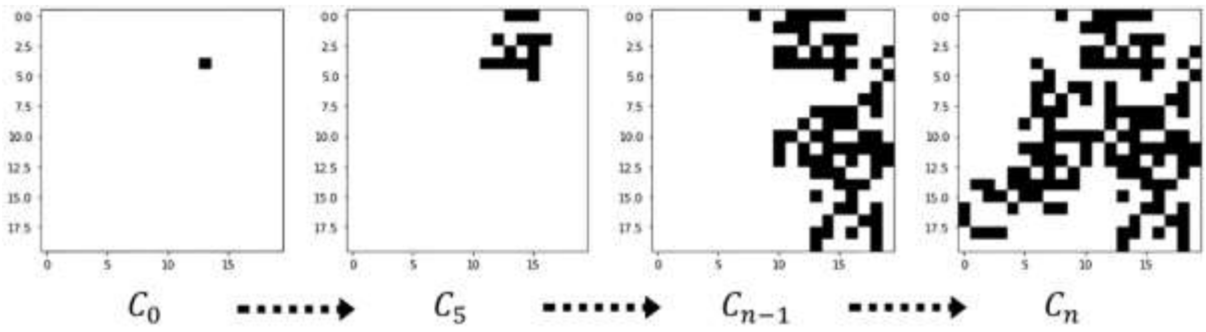


Figure 3. Procedure of agglomerate clustering follows the sequence. The image sequence, $\mathrm{C}_0 \rightarrow \cdots \mathrm{C}_5 \rightarrow \cdots \mathrm{C}_{n-1} \rightarrow \mathrm{C}_n$, is created by applying the agglomerate clustering method to Cluster #5 in Fig. 6d. C_n corresponds to Cluster #5

2.3 Learning Elementary Cellular Automata Rules

ECA is introduced to find rules in the sequences for each cluster. ECA is a one-dimensional array of cells, where each cell takes either 1 or 0, representing BOS or not BOS, respectively. It generates next array depending on its own state and states of its two closest neighbors (Martinez et al. 2012; Weisstein 2017; Wolfram 1983, 2002) . Hence, 256 rules numbering from 0 to 255 are available to represent the sequence evolution. In this study only the even number rules are used. The odd number rules are excluded because it is making the next generation value 1 when both the current cell and the neighboring cells are 0, which is unsuitable for the biological spreading model. Each row in ECA represents one generation, where ECA is a one-dimensional array. The next generation is generated by the ECA rules. By reconstructing a one-dimensional array into a 2D image, each generation can be made of a sequence of images that change according to the ECA rules in Fig. 4.

The rules are learned by training the image change pattern using the Convolutional Neural Network(CNN) (LeCun et al. 2015; Webb 2018). CNNs are a subset of a class of deep learning algorithms, most commonly used for spatial pattern analysis in biology and ecology (Brodrick et al. 2019; Kattenborn et al. 2021; Webb 2018). Additionally, CNN methods can efficiently classify the predicted distributions of many species (Deneu et al. 2021). In this simulation CNNs are trained with Keras package in TensorFlow (Martín et al. 2015).

2.3.1 Generate training data

The procedure is as follows:

• Create a 20 by 20 matrix by random seeding of 1's at 100, 200 and 300 initial points.

• Reshape the 20 by 20 matrix to a 1 by 400 matrix

• Generate the next generation of 1 by 400 matrix according to ECA rules.

• Reshape two consecutive 1D matrices to two consecutive 2D matrices, which are considered as one sets of images, such as $(C_{n-1},\ C_n)$ in Fig. 3.

• Generate sets of image data for all 128 even rules

• Generate 500 sets of image data for each 100, 200 and 300 initial points for each rule

Hence, 500*3*128=192,000 sets of image data are generated

2.3.2 Training the rules

• Separate 80% of training data and 20% of test data from total data

• Learning the rules using CNN(Convolution Neural Network) method (Fig. 4)

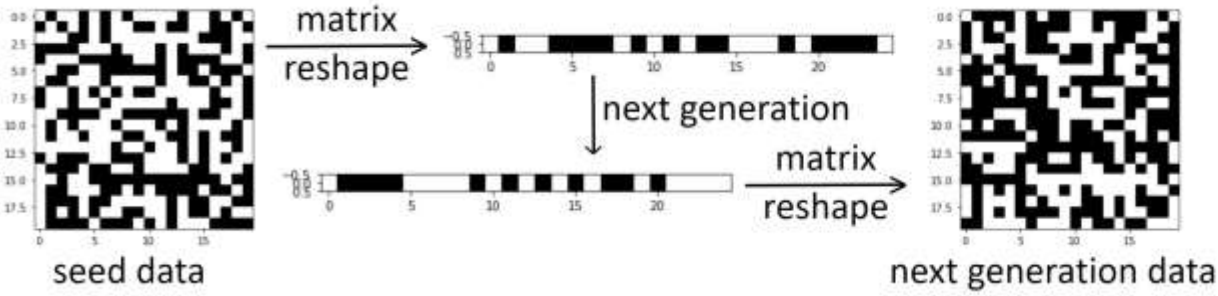


Figure 4. Seed 100, 200 and 300 random points on the 20 by 20 matrix. Reshape 20 by 20 matrix to 1 by 400 matrix, then apply the elementary cellular automata rules to make the new generation of 1 by 400 matrix. Reshape the new matrix to 2 dimensional 20 by 20 matrix. To train the ECA rule learning, we generate 1500 number of matrix pairs for all even rules of ECA by random seeding of 1's at 100, 200 and 300 initial points we generate

2.4 Spreading Intensity

To estimate the intensity of spreading, the expected number of BOS variation depending on the rules governing the evolution of clusters is estimated. As an initial value, a value of 1 is randomly given to 100 cells out of 400 cells of the image, then the number of 1's in the image is counted while evolving over 400 generations according to all even-number rules of ECA. This procesure is repeated 10 times to get the average number of 1's. Figure 5 shows the variaton in the expected number of BOS with the initial value set to 100 and the mean of each expected number of BOS for 400 generations for certain rules. The mean of the expected number of BOS divided by the initial value of 100 is defined as spreading intensity for each rule, which shows the growth rate of the expected number of BOS. The results for all even rules are in Table 3 in Appendix B.

The mean of the expected number of BOS according to each rule is multiplied by the percentile distribution of the rule to get the mean of the expected number of BOS of the cluster. Here the *spreading intensity* is defined as the mean of the number of BOS:

$$\textit{spreading intensity} = \sum_i \text{Percentile of rule } (\mathrm{x}_i) * \text{ mean of the expected number of BOS for rule } (\mathrm{x}_i)/100$$

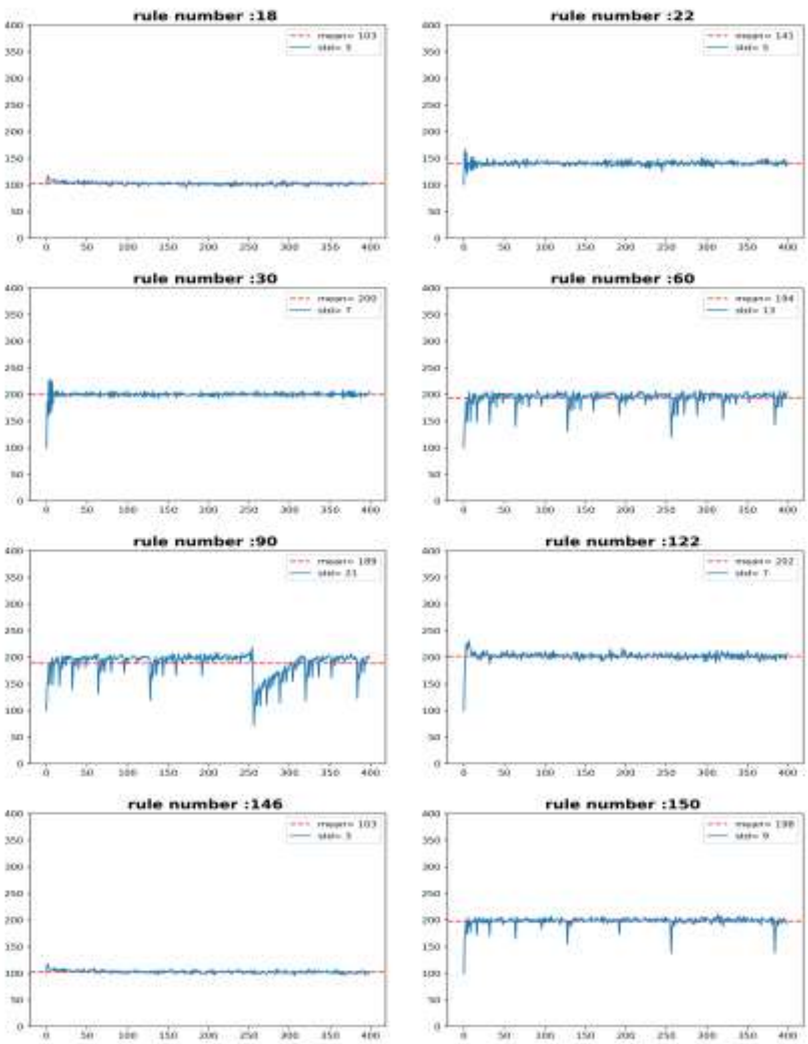


Figure 5. Patterns of the expected number of BOS change by generation according to each rule. A value of 1 is randomly assigned to 100 cells out of 400 cells, and the number of cells having a value of 1 is counted up to 400 generations according to the ECA rule. It shows the variation in the expected number of BOS. Rules 18 and 146 show no significant change over 400 generations, and rules 30, 60, 90, 122, and 150 oscillate around the 200 BOSs, while rule 22 stays around 150 BOSs. The mean of the expected number of BOS is indicated in the legend and shown in red dotted line.

2.5 Assessment of spreading

Since the spreading intensity is evaluated based on the mean of the expected number of BOS only without considering any other environmental and biological variables, the final predicted spreading intensity is weighted by the habitat suitability. The Maxent software (Maximum Entropy, version 3.4.1) is used in estimating the relative habitat suitability of sites by comparing environmental conditions at known observed sites to the available environmental conditions such as precipitation, temperature, elevation and so on (Tesfamariam et al. 2022; Venne and Currie 2021). The main environmental factors when using Maxent software are: annual mean temperature, mean diurnal range, temperature seasonality, annual precipitation, precipitation of wettest month and precipitation of driest month.

3 Simulation Results

3.1 Clustering

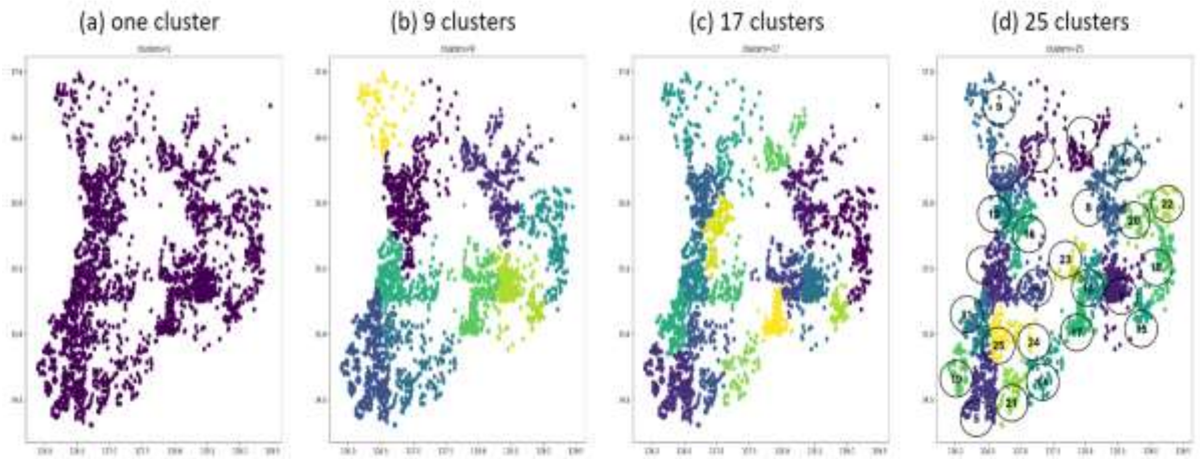


Figure 6. Results of divisive clustering. **a** Observation data **b** divisive clustering after 9 clusters are formed **c** divisive clustering after 17 clusters are formed **d** the size of the clusters became similar to the size of the administrative district at 25 clusters are formed.

Using the hierarchical clustering method, the entire data is divided into 25 small clusters, and the size of the clusters became similar to the size of the administrative district (Fig. 6d). The number of clusters can be set to 1, 9 or 17, depending on the size of the region of interest (Fig. 6a-c). The common feature of the clusters is the high density mainly around the waterside and wetland. However, the shape of the cluster alone doesn't represent the spreading intensity for each cluster. Biological and environmental informations are not taken into account when grouping the clusters

3.2 Learning the rules Using CNN

Figure 7 shows the accuracy of training the rules of ECA. The accuracy is more than 99%. This would mean that the rule of changes in the bullfrog distribution could be learned with a very high confidence.

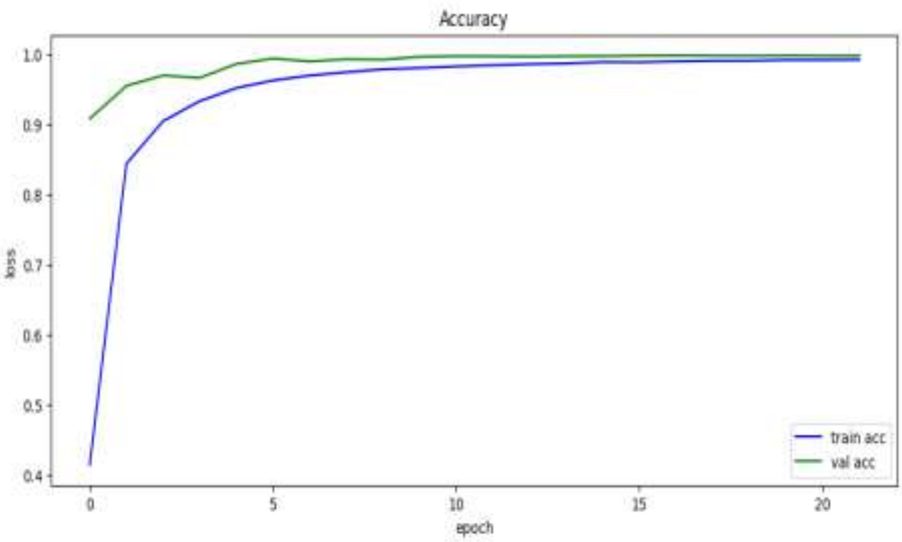


Figure 7. *The accuracy of training. The x-axis (epoch) represents the number of training iterations, and the y-axis (accuracy) represents the accuracy of the machine learning (the blue line (train acc; train accuracy) is the accuracy curve of the training data. The green line (val acc; validate accuracy) represents is the accuracy curve of the validation data). CNN is trained to learn ECA rules*

3.3 Spreading intensity

Figure 8 in Appendix A shows the distribution of rules predicted through CNN learning for each cluster in Fig. 6d. The expected number of BOS for all 128 ECA rules estimated over 400 generations are in Table 3 in Appendix B. For cluster 14 as an example, it shows a distribution of 84.5% for rule 204, 8.5% for 206, and 7.0% for rule 220 in Fig. 9. The mean of the convergent number for rule 204 is 100, for the rule 206 is 323, and for the rule 220 is 322. Therefore, if bullfrogs are found in 100 cells now, the expected number of converged BOS in cluster 14 can be calculated as 1.00*0.845+3.23*0.085+3.22*0.07=1.34495, which is the spreading intensity for the cluster 14. The spreading intensity of all clusters are shown in Table 1

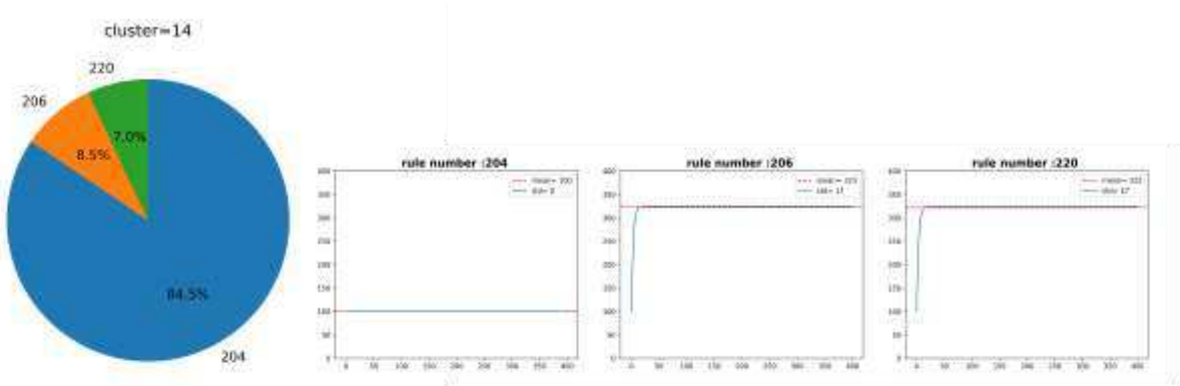


Figure 9. Distribution of rules for cluster 14. It consists of rules 204, 206, and 220 corresponding to 84.5%, 8.5%, and 7.0% respectively. The mean of the convergent number of BOS for rule 204 is 100, for the rule 206 is 323, and for the rule 220 is 322

3.4 Spreading Assessment

Figure 10 shows the Spreading Intensity (SI), Habitat Suitability (HS) and Spreading Assessment (SA) of 25 clusters. Figure 10b shows the SI distribution. It does not reflect environmental and biological variables, and it shows the spreading intensity calculated only by machine learning methods (clustering, CNN, etc.). Areas that are already saturated may have low SI values, and areas with low saturation, such as mountainous areas, may have large SI values. Figure 10c represents HS distributions. HS values obtained by using Maxent software reflect ecological environmental factors for the bullfrogs. Figure 10d shows the distribution of SA values obtained by multiplying SI values and HS values. The HS value ranges from 0 to 1, and the closer it is to 1, the more suitable. All distribution values are relative and expressed in four stages: strong spreading, weak spreading, strong retention, and weak retention.

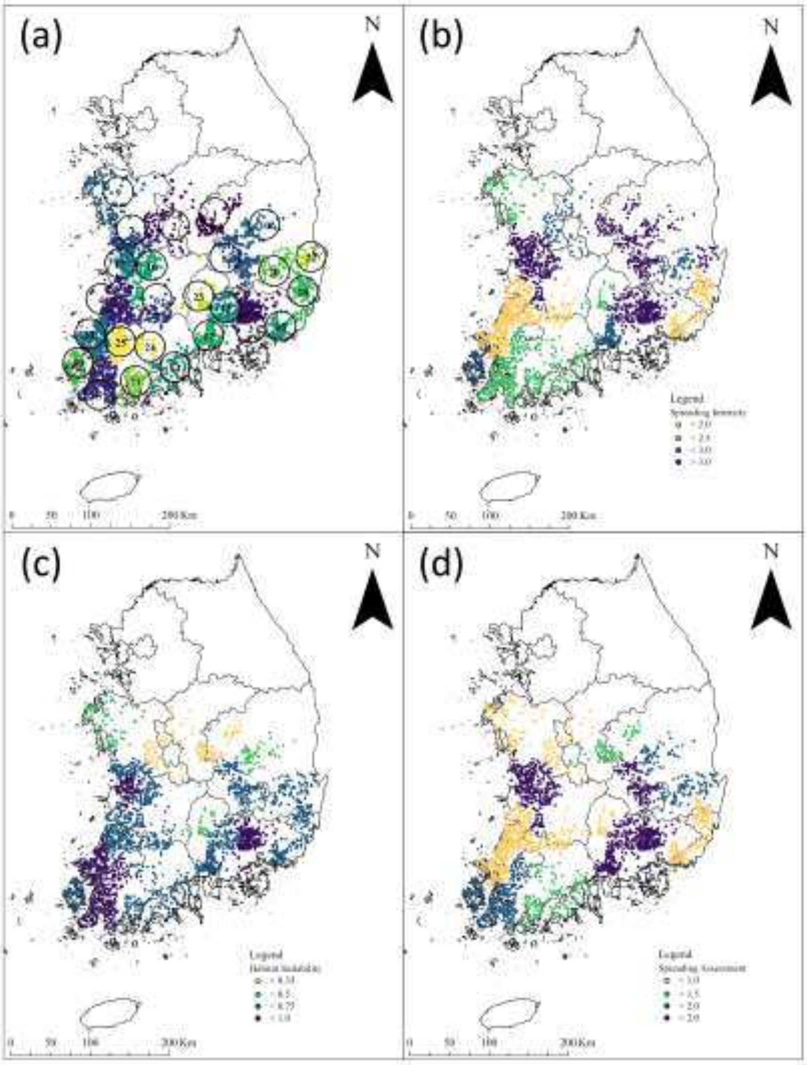


Figure 10. Results of 25 clusters. **a** 25 Clusters : Divisive clustering is performed until the clustering became similar to the local administrative districts. **b** Spreading Intensity(SI): It does not show a strong spreading intensity in coastal and wetland areas. This suggests the possibility that spreading has already occurred to saturation. **c** Habitat Suitability(HS): Habitat suitability calculated using Maxent software. If the SI, the spreading intensity, is weak at a high HS, it means that spreading has already occurred sufficiently. **d** Spreading Assessment(SA): Areas with high probability of spreading are marked with red dots

Table 1 shows the result of calculating the spreading assessment. From left to right, each column represents the number of BOS per cluster, spreading intensity, habitat suitability, spreading assesment, and geometric center latitude and longitude. The higher the value, the greater the probability of spreading. SI is a value obtained through machine learning using only BOS data. Environmental and biological factors were reflected through Habitat Suitability (HS) to get Spreading Assessment (SA).

Table 1 The results of bullfrog spreading for 25 clusters

Clustering Number	number of BOS	SI	HS	SA	Longitude	Latitude
1	77	3.25	0.331892778	1.078651528	127.99	36.42
2	83	2.87	0.336620682	0.966101357	127.27	36.42
3	267	3.1	0.772864513	2.395879991	128.47	35.36
4	202	1	0.740367631	0.740367631	126.64	35.47
5	168	2.48	0.794083093	1.96932607	126.43	34.58
6	103	1.51	0.594828446	0.898190953	127.2	35.36
7	127	3.25	0.713751758	2.319693214	126.77	36.16
8	125	3.24	0.65006152	2.106199325	128.33	36
9	96	2.08	0.460205108	0.957226626	126.47	36.71
10	58	3.46	0.473968133	1.639929741	128.6	36.33
11	205	1.26	0.791182732	0.996890243	126.46	35.08
12	135	3.25	0.781342992	2.539364724	126.76	35.93
13	98	3.42	0.62174352	2.12636284	128.14	35.4
14	68	2.23	0.634174926	1.414210085	127.36	34.73
15	87	1.35	0.690117529	0.931658665	128.88	35.18
16	113	3.16	0.673967381	2.129736922	127.01	35.78
17	136	2.91	0.701436489	2.041180182	127.91	35.11
18	107	1.1	0.62500829	0.687509118	129.21	35.57
19	70	2.66	0.697356021	1.854967015	126.11	34.73
20	65	2.86	0.6193504	1.771342144	128.88	35.89
21	58	2.14	0.617724528	1.321930491	126.88	34.58
22	30	3.33	0.573416667	1.9094775	129.28	35.99
23	37	2.02	0.3649146	0.737127492	127.81	35.53
24	37	2.21	0.575529162	1.271919448	127.03	34.99
25	65	2.22	0.804474672	1.785933772	126.66	34.98

The final spreading assessment is the spreading intensity multiplied by the habitat suitability estimated by Maxent software 3.4.1. The higher the value, the greater the probability of spreading. Table 2 shows relative spreading assessments. Four cluster groups are created based on assessment scores. The clusters in groups (I) and (II) show spreading assessment scores greater than 2, which means that they will continue to spread. Clusters in group (III) show the scores of 1 to 1.5, which can be considered as slow spreading or maintaining the population. For group (IV) clusters, the spread appears to have stopped, and the population may decline, especially in clusters #4, #8, and #23

Table 2 Groups of Spreading Assessment

Group	Spreading Assessment (SA)	Cluster Number(#)	Relative Results
Group (I)	2.0 <SA	3,7,8,12,13,16,17	Continue to spread intensively
Group (II)	1.5<SA<2.0	5,10,19,20,22,25	Continue to spread
Group (III)	1<SA<1.5	1,14,21,24	maintain population
Group (IV)	SA<1	2,4,6,9,11,15,18,23	maintain population and possibly decrease in 4,18,23

4 Discussion

Clustering is performed until 25 clusters are formed to roughly match the size of administrative districts. If the stopping number of clusters is changed in this method, the target range and the convergence of the expected number of BOS may also change, so the number of clusters should be adjusted to properly include the region of interest.

In this study, the numpy. reshape() function (Harris et al. 2020) was used to rearrange two-dimensional images into one-dimensional array. Another future study is needed to apply ECA according to various array arrangements. When applying the ECA rules, zero padding was applied to both end points, that is, 0 is used for the -1th and 401st virtual cells. It is assumed that Bullfrog has been never found outside the cluster. If found, they should be included in other clusters.

Since the agglomerate clustering method is used in generating image sequences to train convolutional neural network, rules for increasing the expected number of BOS were mainly distributed as shown in Fig. 8 in Appendix A. However in some rules the expected number of BOS appeared to increase rapidly at first show various patterns, such as converging to a certain value (e.g. rule 124) or decreasing over time (e.g. rule 120) in Table 3 in Appendix B.

In estimating spreading intensity, the mean of the expexted number of BOS is used, but the slope can be more useful in expressing the tendency of spreading as shown in Table 3 in Appendix B. Further research is needed to define the appropriate diffusion strength according to the convergence pattern of the expected number of BOS.

Since the spreading intensity is estimated based only on the data currently found it is relatively low in the region where spreading is already completed. Low spreading intensity may mean that it is already saturated, which is different from extinction. Alternatively, the carrying capacity may decrease from a population dynamics perspective due to the emergence of natural enemies or human quarantine.

The size of cells is identical in each cluster. The density of observations in a cell is uniformly 0 or 1, where 1 being BOS. However, in each cell, the number of bullfrog observations are different. Hence, the cloglog(default in Maxent 3.4.1) option is used to treat occurrence records as points rather than grid cells to estimate relative habitat suitability (Steven J Phillips et al. 2017).

Geographical characteristics and ecological characteristics are replaced by habitat suitability using Maxent but more detailed cultural characteristics should be applied. In addition to observations, appropriate detection methods for bullfrogs, such as eDNA method~~s~~ or audio recording devices, are required (Kamoroff et al. 2020).

Only the accuracy of machine learning is presented as a verification method. In order to verify its validity, it is necessary to select 3 or 4 regions and observe the spreading intensity continuously for several years to generate time series data and compare it with the expected values from simulations.

5 Conclusion

In this paper, the relative spreading assessment is estimated for South Korea using machine learning methods. The extent to which bullfrogs continue to spread at observation sites is quantified and assessed.

Since there is no time series data, the accumulated data are used to evaluate the spread of bullfrogs by creating a spatial series using machine learning. In this process, biological and environmental factors were not considered at all. Habitat suitability obtained by using Maxent software includes environmental and biological factors, which were applied in the form of weights to the final spread classification evaluation

The cell where bullfrogs are found (BOS: bullfrog observed site) is assigned to 1, and the number of 1's in 400 cells composed of 1's and 0's is counted and used as the spreading index of bullfrogs. The mean of the number of BOS divided by the initial value of 100 is assumed as a measure of spreading intensity for each rule. The spreading intensity is weighted by the percentile of the rules estimated by the CNN method. Spreading assessment is the spreading intensity multiplied by the habitat suitability, which can be used as one of the index showing the tendency of spreading.

Acknowledgements This work was supported by Korea Environment Industry & Technology Institute(KEITI) through Exotic Invasive Species Management Program, funded by Korea Ministry of Environment(MOE)(2018002270001). Additional funding was provided by the National Research Foundation of Korea (NRF), which is funded by the Ministry of Education, Republic of Korea (NRF-2020R1I1A3071769).

Authors' contributions: Conceptualization and methodology, G.O., H.S. and H.J.; software and validation, G.O., Y.W., H.K., S.C. and H.J.; formal analysis, investigation and resources, Y.W.,H.K., Y.K., S.C. and H.J.; writing—original draft preparation, G.O. and H.J.; writing—review and editing, Y.K., Y.W. and H.J.; visualization, G.O. and H.J.; supervision, H.S. and H.J.; funding acquisition, H.S. and H.J.

Correspondence and requests for materials should be addressed to H.J.

Data Availibility All data generated or analysed during this study are included in this published article.

A Percentile Distribution of probabilities of predicted rules for 25 clusters

The trained CNN gives us the rule distribution for 25 clusters. In fact, the CNN provides a percentile distribution for 128 rules, of which only those that occur more than 1% are shown in the circle diagram.

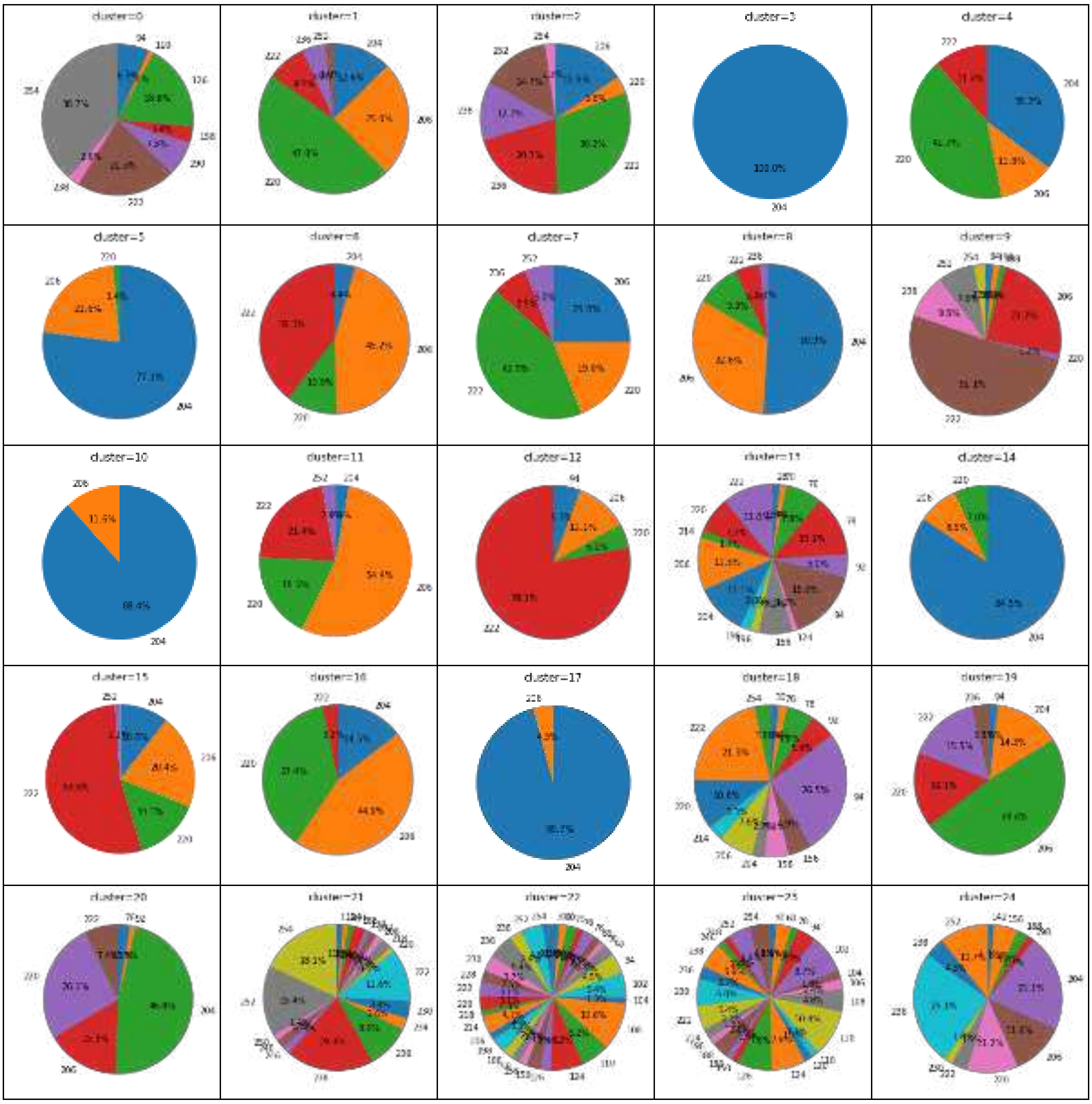
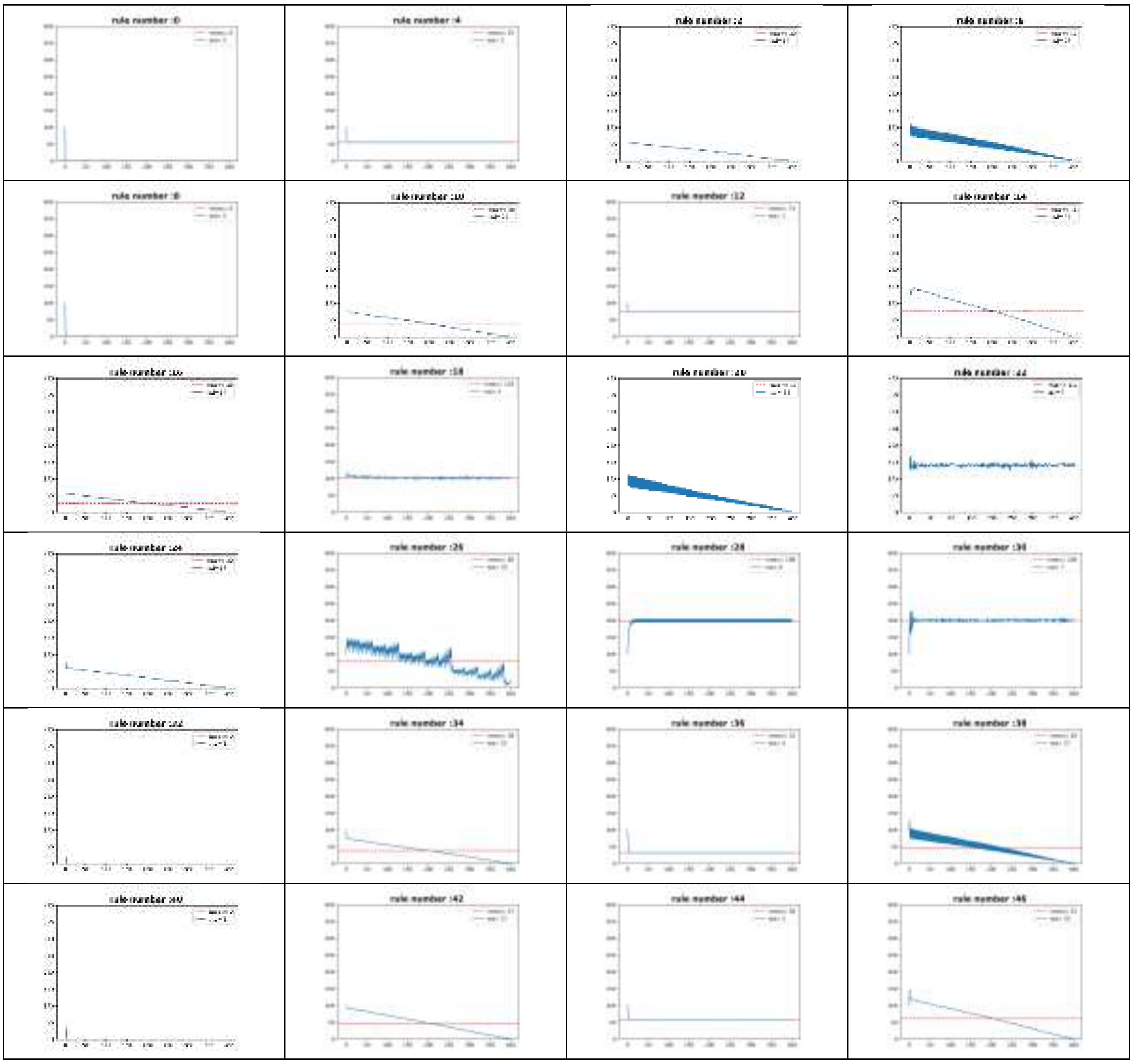


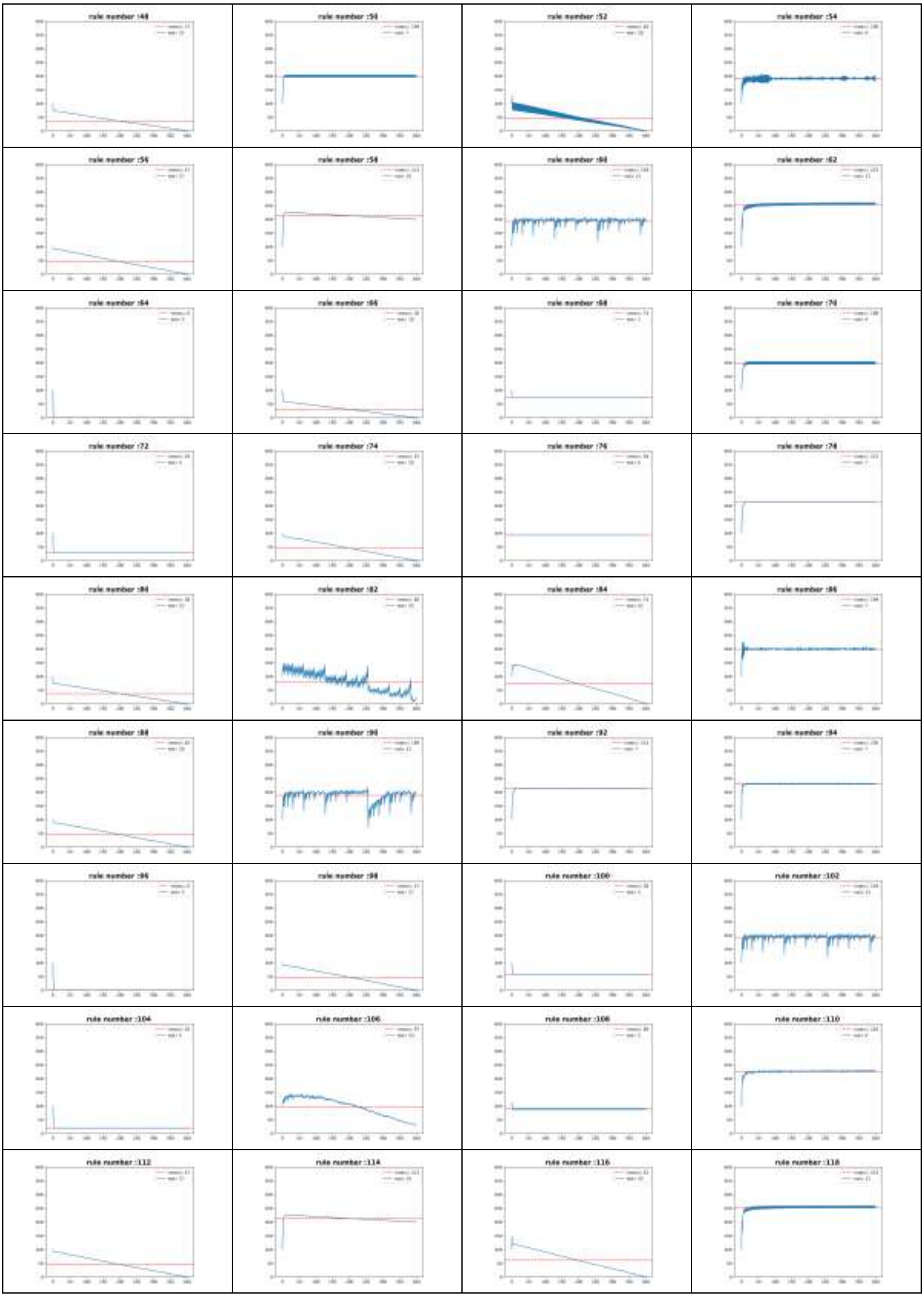
Figure 8. Percentile Distribution of predicted rules by cluster. It shows percentile distribution of the rules for 25 clusters. It is obtained by learning ECA rules to each cluster. The percentile distribution of the rules is obtained using the trained parameters of CNN.

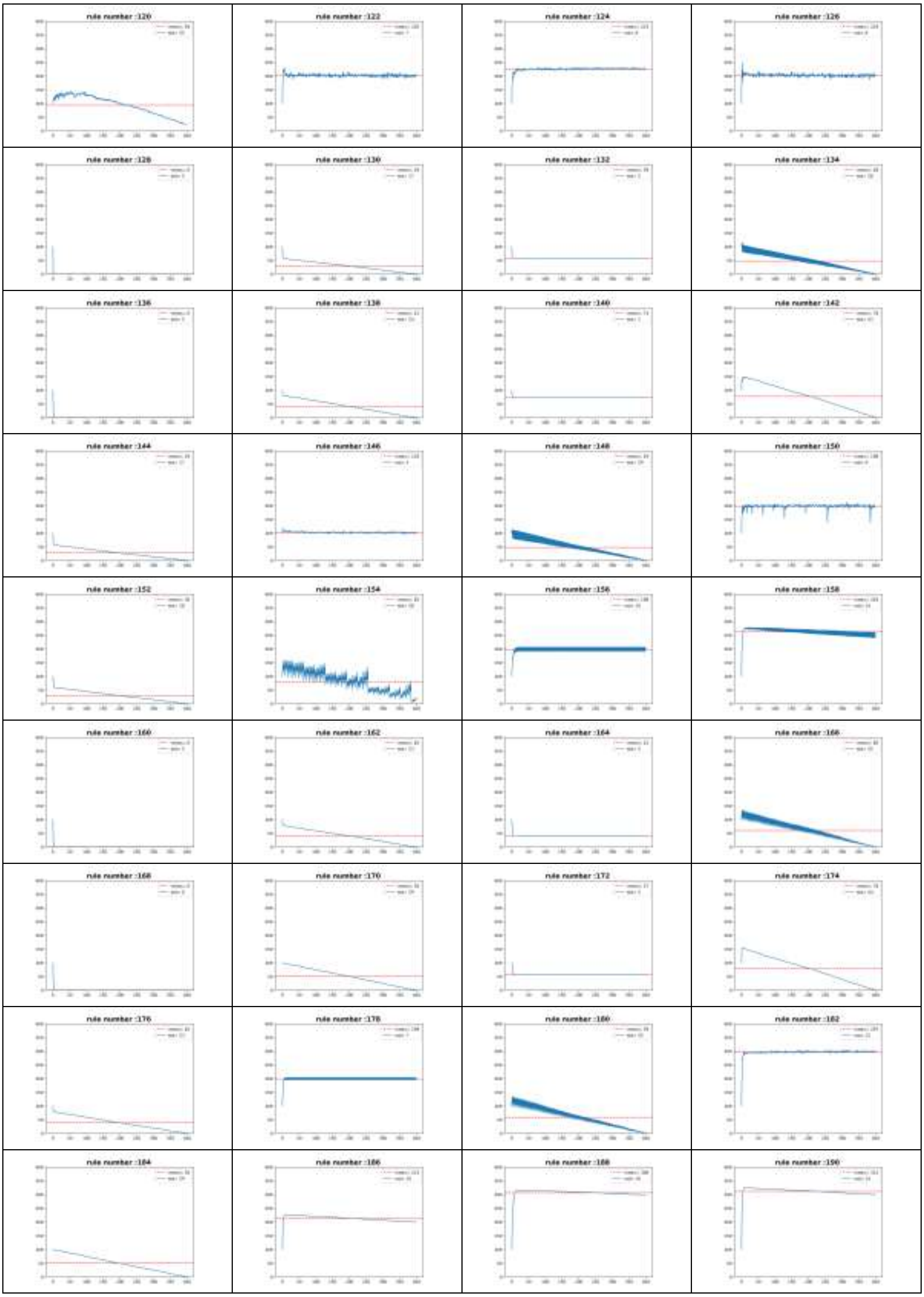
B Variation patterns of even ECA rules.

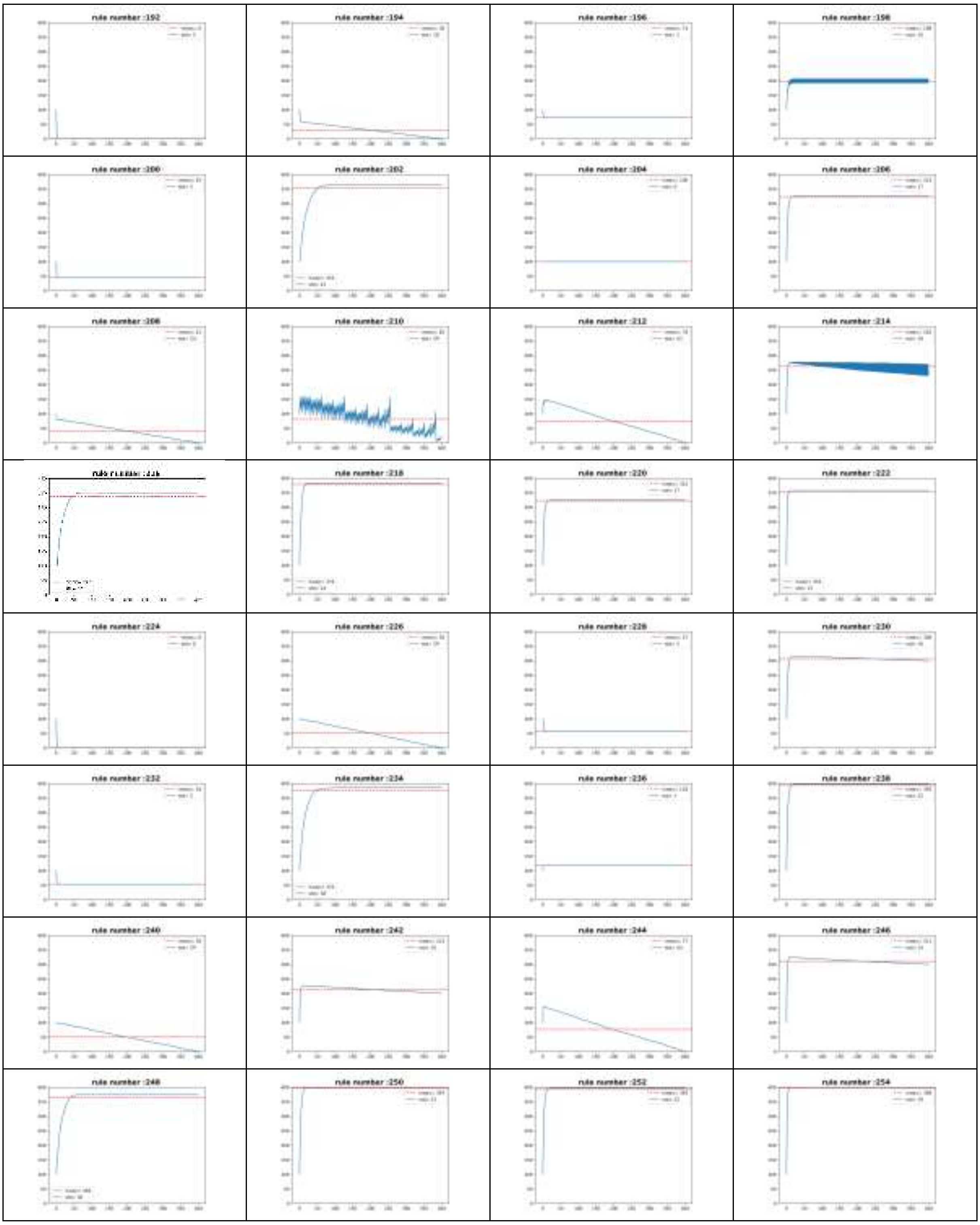
A value of 1 is randomly assigned to 100 cells out of 400 cells, and the number of cells having a value of 1 is counted up to 400 generations according to ECA rules. This process is repeated 10 times and the average value is obtained. Cells with a value of 1 correspond to the Bullfrog Observed Sites (BOS). The mean value is shown as a red dotted line, and according to the rules, the number of estimated BOS either maintains the mean value line, decreases, or shows various types of oscillations around the line.

Table 3. Variations in the expected number of BOS over 400 generations according to the ECA even number rules.




rule number :48
rule number :50
rule number :52
rule number :54
rule number :56
rule number :58
rule number :60
rule number :62
rule number :64
rule number :66
rule number :68
rule number :70
rule number :72
rule number :74
rule number :76
rule number :78
rule number :80
rule number :82
rule number :84
rule number :86
rule number :88
rule number :90
rule number :92
rule number :94
rule number :96
rule number :98
rule number :100
rule number :102
rule number :104
rule number :106
rule number :108
rule number :110
rule number :112
rule number :114
rule number :116
rule number :118


rule number :120
rule number :122
rule number :124
rule number :126
rule number :128
rule number :130
rule number :132
rule number :134
rule number :136
rule number :138
rule number :140
rule number :142
rule number :144
rule number :146
rule number :148
rule number :150
rule number :152
rule number :154
rule number :156
rule number :158
rule number :160
rule number :162
rule number :164
rule number :166
rule number :168
rule number :170
rule number :172
rule number :174
rule number :176
rule number :178
rule number :180
rule number :182
rule number :184
rule number :186
rule number :188
rule number :190


rule number :192
rule number :194
rule number :196
rule number :198
rule number :200
rule number :202
rule number :204
rule number :206
rule number :208
rule number :210
rule number :212
rule number :214
rule number :216
rule number :218
rule number :220
rule number :222
rule number :224
rule number :226
rule number :228
rule number :230
rule number :232
rule number :234
rule number :236
rule number :238
rule number :240
rule number :242
rule number :244
rule number :246
rule number :248
rule number :250
rule number :252
rule number :254

Reference

Brodrick Philip G, Davies Andrew B, Asner Gregory P (2019) Uncovering ecological patterns with convolutional neural networks. Trends in ecology & evolution 34(8): 734-745. https://doi.org/10.1016/j.tree.2019.03.006

Chang Byungwoo, Kim Inyoo, Choi Kwanghun, Cho Wonhee, Ko Dongwook W (2022) Population Dynamics of American Bullfrog (Lithobates catesbeianus) and Implications for Control. Animals 12(20): 2827.

da Silveira Vasconcelos Tiago, Rodríguez Miguel Ángel, Hawkins Bradford Alan (2011) Biogeographic distribution patterns of South American amphibians: a regionalization based on cluster analysis. Journal of Biogeography 39: 1720-1732. https://doi.org/10.4322/natcon.2011.008

Database Global Invasive Species (2023) Global Invasive Species Database. http://www.iucngisd.org/gisd/100_worst.php. Accessed 28 Mar 2023

Deneu Benjamin, Servajean Maximilien, Bonnet Pierre, Botella Christophe, Munoz François, Joly Alexis (2021) Convolutional neural networks improve species distribution modelling by capturing the spatial structure of the environment. PLoS computational biology 17(4): e1008856. https://doi.org/10.1371/journal.pcbi.1008856

Duryea Jack (2018) Cellular Automata and Their Elegant Complexities: Musings about Blockchain, Quantum Computing, and AI. Medium HackerNoon.com. https://medium.com/hackernoon/cellular-automata-and-their-elegant-complexities-musings-about-blockchain-quantum-computing-and-c8d2c2d1cf70. Accessed 15 Jul 2007

Elith* Jane, H. Graham* Catherine, P. Anderson Robert, Dudík Miroslav, Ferrier Simon, Guisan Antoine, et al. (2006) Novel methods improve prediction of species' distributions from occurrence data. Ecography 29(2): 129-151. https://doi.org/10.1111/j.2006.0906-7590.04596.x

Ermentrout G Bard, Edelstein-Keshet Leah (1993) Cellular automata approaches to biological modeling. Journal of theoretical biology 160(1): 97-133. https://doi.org/10.1006/jtbi.1993.1007

Ficetola Gentile Francesco, Maiorano Luigi, Falcucci Alessandra, Dendoncker Nicolas, Boitani Luigi, PADOA-SCHIOPPA EMILIO, et al. (2010) Knowing the past to predict the future: land-use change and the distribution of invasive bullfrogs. Global change biology 16(2): 528-537. https://doi.org/10.1111/j.1365-2486.2009.01957.x

Ficetola Gentile Francesco, Thuiller Wilfried, Miaud Claude (2007) Prediction and validation of the potential global distribution of a problematic alien invasive species—the American bullfrog. Diversity and distributions 13(4): 476-485. https://doi.org/10.1111/j.1472-4642.2007.00377.x

Giovanelli Joao GR, Haddad Célio FB, Alexandrino Joao (2008) Predicting the potential distribution of the alien invasive American bullfrog (Lithobates catesbeianus) in Brazil. Biological Invasions 10: 585-590. https://doi.org/10.1007/s10530-007-9154-5

Groffen Jordy, Kong Sungsik, Jang Yikweon, Borzée Amaël (2019) The invasive American bullfrog (Lithobates catesbeianus) in the Republic of Korea: history and recommendations for population control. Management of Biological Invasions 10(3): 517. http://dx.doi.org/10.3391/mbi.2019.10.3.08

Harris Charles R, Millman K Jarrod, Van Der Walt Stéfan J, Gommers Ralf, Virtanen Pauli, Cournapeau David, et al. (2020) Array programming with NumPy. nature 585(7825): 357-362.

Iñiguez Carlos A, Morejón Felipe J (2012) Potential distribution of the American bullfrog (Lithobates catesbeianus) in Ecuador. south american Journal of herpetology 7(2): 85-90. https://doi.org/10.2994/057.007.0211

Jang HJ, Suh JH (2010) Distribution of amphibian species in South Korea. Korean Journal of Herpetology 2: 45-51.

Kamoroff Colleen, Daniele Ninette, Grasso Robert L, Rising Rebecca, Espinoza Travis, Goldberg Caren S (2020) Effective removal of the American bullfrog (Lithobates catesbeianus) on a landscape level: long term monitoring and removal efforts in Yosemite Valley, Yosemite National Park. Biological Invasions 22(2): 617-626. https://doi.org/10.1007/s10530-019-02116-4

Kang Hee-Jin, Koo Kyo Soung, Sung Ha-Cheol (2019) Current distribution of American bullfrog Rana catesbeiana Shaw, 1802 in the Republic of Korea. BioInvasions Records 8(4): 942-946. https://doi.org/10.3391/bir.2019.8.4.23

Kattenborn Teja, Leitloff Jens, Schiefer Felix, Hinz Stefan (2021) Review on Convolutional Neural Networks (CNN) in vegetation remote sensing. ISPRS journal of photogrammetry and remote sensing 173: 24-49. https://doi.org/10.1016/j.isprsjprs.2020.12.010

Kim JB (2009) Taxonomic list and distribution of Korean amphibians. Korean J. Herpetol 1: 1-13.

Koo Kyo Soung, Choe Minjee (2021) Distribution Change of Invasive American Bullfrogs (Lithobates catesbeianus) by Future Climate Threaten Endangered Suweon Treefrog (Hyla suweonensis) in South Korea. Animals 11(10): 2865. https://doi.org/10.3390/ani11102865

LeCun Yann, Bengio Yoshua, Hinton Geoffrey (2015) Deep learning. nature 521(7553): 436-444. https://doi.org/10.1038/nature14539

Martín Abadi, Ashish Agarwal, Paul Barham, Eugene Brevdo, Zhifeng Chen, Craig Citro, et al. (2015) TensorFlow: Large-Scale Machine Learning on Heterogeneous Systems. https://www.tensorflow.org/. Accessed

Martinez Genaro J, Adamatzky Andrew, Alonso-Sanz Ramon (2012) Complex dynamics of elementary cellular automata emerging from chaotic rules. International Journal of Bifurcation and Chaos 22(02): 1250023. https://doi.org/10.1142/S021812741250023X

Nagatani Takashi, Tainaka Kei-ichi (2018) Cellular automaton for migration in ecosystem: Application of traffic model to a predator–prey system. Physica A: Statistical Mechanics and its Applications 490: 803-807. https://doi.org/10.1016/j.physa.2017.08.151

No Sun-Ho, Jung Jin-Seok, You Young-Han (2017) Ecological control of invasive alien species, American bullfrog (Rana catesbeiana) using native predatory species. Korean Journal of Environment and Ecology 31(1): 54-61. https://doi.org/10.13047/KJEE.2017.31.1.054

Nori Javier, Urbina-Cardona J Nicolás, Loyola Rafael D, Lescano Julian N, Leynaud Gerardo C (2011) Climate change and American bullfrog invasion: what could we expect in South America? PloS one 6(10): e25718. https://doi.org/10.1371/journal.pone.0025718

Oh Hong-Shik, Hong Chang-Eui (2007) Current conditions of habitat for Rana catesbeiana and Trachemys scripta elegans imported to Jeju-do, including proposed management plans. Korean Journal of Environment and Ecology 21(4): 311-317.

Park Hye-Rin, Rahman Md Mizanur, Park Seung-Min, Choi Jae-Hyeok, Kang Hee-Jin, Sung Ha-Cheol (2022) Risk assessment for the native anurans from an alien invasive species, American bullfrogs (Lithobates catesbeianus), in South Korea. Scientific reports 12(1): 13143. https://doi.org/10.1038/s41598-022-17226-8

Patlolla Chaitanya Reddy. (2018). Understanding the concept of Hierarchical clustering Technique. Retrieved from Medium website: https://towardsdatascience.com/understanding-the-concept-of-hierarchical-clustering-technique-c6e8243758ec

Pedregosa Fabian, Varoquaux Gaël, Gramfort Alexandre, Michel Vincent, Thirion Bertrand, Grisel Olivier, et al. (2011) Scikit-learn: Machine learning in Python. the Journal of machine Learning research 12: 2825-2830.

Phillips Steven J, Anderson Robert P, Dudík Miroslav, Schapire Robert E, Blair Mary E (2017) Opening the black box: An open-source release of Maxent. Ecography 40(7): 887-893. https://doi.org/10.1111/ecog.03049

Phillips Steven J, Anderson Robert P, Schapire Robert E (2006) Maximum entropy modeling of species geographic distributions. Ecological modelling 190(3-4): 231-259. https://doi.org/10.1016/j.ecolmodel.2005.03.026

Phillips Steven J, Dudík Miroslav (2008) Modeling of species distributions with Maxent: new extensions and a comprehensive evaluation. Ecography 31(2): 161-175. https://doi.org/10.1111/j.0906-7590.2008.5203.x

Phillips Steven J. (2017) A Brief Tutorial on Maxent. http://biodiversityinformatics.amnh.org/open_source/maxent/. Accessed 14 Aug 2023

Qin Jiaohua, Pan Wenyan, Xiang Xuyu, Tan Yun, Hou Guimin (2020) A biological image classification method based on improved CNN. Ecological Informatics 58: 101093. https://doi.org/10.1016/j.ecoinf.2020.101093

Schlaepfer Martin A, Sherman Paul W, Blossey Bernd, Runge Michael C (2005) Introduced species as evolutionary traps. Ecology Letters 8(3): 241-246. https://doi.org/10.1111/j.1461-0248.2005.00730.x

Tesfamariam Birhane Gebrehiwot, Gessesse Berhan, Melgani Farid (2022) MaxEnt-based modeling of suitable habitat for rehabilitation of Podocarpus forest at landscape-scale. Environmental

Systems Research 11(1): 1-12. https://doi.org/10.1186/s40068-022-00248-6

Venne Simon, Currie David J (2021) Can habitat suitability estimated from MaxEnt predict colonizations and extinctions? Diversity and distributions 27(5): 873-886. https://doi.org/10.1111/ddi.13238

Webb Sarah (2018) Deep learning for biology. nature 554(7693): 555-557. https://doi.org/10.1038/d41586-018-02174-z

Weisstein Eric W (2017) Elementary Cellular Automaton. MathWorld--A Wolfram Web Resource. https://mathworld.wolfram.com/ElementaryCellularAutomaton.html. Accessed

Wolfram Stephen (1983) Statistical mechanics of cellular automata. Reviews of modern physics 55(3): 601. https://doi.org/10.1103/RevModPhys.55.601

Wolfram Stephen (2002) *A new kind of science* (Vol. 5). Wolfram media Champaign, IL.