

Presence of Amphibian Species Prediction Using Features Obtained from GIS and Satellite Images

Nadia Shaker Habib, Omar Kamal Abu Maghasib, Ahmed Rashad Al-Ghazali, Bassem S. Abu-Nasser, Samy S. Abu-Naser

Department of Information Technology,
Faculty of Engineering and Information Technology,
Al-Azhar University - Gaza, Palestine

Abstract: *The establishment of the transport infrastructure is usually preceded by an EIA procedure, which should determine amphibian breeding sites and migration routes. However, evaluation is very difficult due to the large number of habitats spread over a vast area and the limited time available for field work. An artificial Neural Network (ANN) is proposed for predicting the presence of amphibians species near the water reservoirs based on features obtained from GIS systems and satellite images. The dataset collected from UCI Machine Learning repository. The dataset is a multi-label classification problem. The goal of this study is to predict the presence of amphibians species near the water reservoirs based on features obtained from GIS systems and satellite images. After preprocessing the data, the proposed model was trained and evaluated. The accuracy of the proposed model for predicting the presence of amphibian's species was 100%.*

Keywords: ANN, JNN, amphibian, multi-label classification,

1. Introduction

Amphibians are ectothermic, tetrapod vertebrates of the class Amphibia. All living amphibians belong to the group Lissamphibia. They inhabit a wide variety of habitats, with most species living with in terrestrial, fossorial, arboreal or freshwater aquatic ecosystems. Thus amphibians typically start out as larvae living in water, but some species have developed behavioral adaptations to bypass this.

The young generally undergo metamorphosis from larva with gills to an adult air-breathing form with lungs. Amphibians use their skin as a secondary respiratory surface and some small terrestrial salamanders and frogs lack lungs and rely entirely on their skin. They are superficially similar to lizards but, along with mammals and birds, reptiles are amniotes and do not require water bodies in which to breed. With their complex reproductive needs and permeable skins, amphibians are often ecological indicators; in recent decades there has been a dramatic decline in amphibian populations for many species around the globe [1].

The earliest amphibians evolved in the Devonian period from sarcopterygian fish with lungs and bony-limbed fins, features that were helpful in adapting to dry land. They diversified and became dominant during the Carboniferous and Permian periods, but were later displaced by reptiles and other vertebrates. Over time, amphibians shrank in size and decreased in diversity, leaving only the modern subclass Lissamphibia.

The three modern orders of amphibians are Anura (the frogs and toads), Urodela (the salamanders), and Apoda (the caecilians). The number of known amphibian species is approximately 8,000, of which nearly 90% are frogs. The smallest amphibian (and vertebrate) in the world is a frog from New Guinea (*Paedophryne amauensis*) with a length of just 7.7 mm (0.30 in). The largest living amphibian is the 1.8 m (5 ft 11 in) South China giant salamander (*Andrias sligoi*), but this is dwarfed by the extinct 9 m (30 ft) *Prionosuchus* from the middle Permian of Brazil. The study of amphibians is called batrachology, while the study of both reptiles and amphibians is called herpetology [2].

The word "amphibian" is derived from the Ancient Greek term ἀμφίβιος (*amphíbios*), which means "both kinds of life", ἀμφί meaning "of both kinds" and βίος meaning "life". The term was initially used as a general adjective for animals that could live on land or in water, including seals and otters. Traditionally, the class Amphibia includes all tetrapod vertebrates that are not amniotes. Amphibia in its widest sense (*sensu lato*) was divided into three subclasses, two of which are extinct [3:]

- Subclass Lepospondyli† (small Paleozoic group, which are more closely related to amniotes than Lissamphibia)
- Subclass Temnospondyli† (diverse Paleozoic and early Mesozoic grade)
- Subclass Lissamphibia (all modern amphibians, including frogs, toads, salamanders, newts and caecilians)
 - Salientia (frogs, toads and relatives): Jurassic to present—6,200 current species in 53 families
 - Caudata (salamanders, newts and relatives): Jurassic to present—652 current species in 9 families
 - Gymnophiona (caecilians and relatives): Jurassic to present—192 current species in 10 families

The actual number of species in each group depends on the taxonomic classification followed. The two most common systems are the classification adopted by the website AmphibiaWeb, University of California, Berkeley and the classification by herpetologist Darrel Frost and the American Museum of Natural History, available as the online reference database "Amphibian Species of the

World". The numbers of species cited above follows Frost and the total number of known amphibian species as of March 31, 2019 is exactly 8,000, of which nearly 90% are frogs [3].

With the phylogenetic classification, the taxon Labyrinthodontia has been discarded as it is a polyparaphyletic group without unique defining features apart from shared primitive characteristics. Classification varies according to the preferred phylogeny of the author and whether they use a stem-based or a node-based classification. Traditionally, amphibians as a class are defined as all tetrapods with a larval stage, while the group that includes the common ancestors of all living amphibians (frogs, salamanders and caecilians) and all their descendants is called Lissamphibia. The phylogeny of Paleozoic amphibians is uncertain, and Lissamphibia may possibly fall within extinct groups, like the Temnospondyli (traditionally placed in the subclass Labyrinthodontia) or the Lepospondyli, and in some analyses even in the amniotes. This means that advocates of phylogenetic nomenclature have removed a large number of basal Devonian and Carboniferous amphibian-type tetrapod groups that were formerly placed in Amphibia in Linnaean taxonomy, and included them elsewhere under cladistic taxonomy. If the common ancestor of amphibians and amniotes is included in Amphibia, it becomes a paraphyletic group [4].

All modern amphibians are included in the subclass Lissamphibia, which is usually considered a clade, a group of species that have evolved from a common ancestor. The three modern orders are Anura (the frogs and toads), Caudata (or Urodela, the salamanders), and Gymnophiona (or Apoda, the caecilians). It has been suggested that salamanders arose separately from a Temnospondyl-like ancestor, and even that caecilians are the sister group of the advanced reptiliomorph amphibians, and thus of amniotes. Although the fossils of several older proto-frogs with primitive characteristics are known, the oldest "true frog" is *Prosalirus bitis*, from the Early Jurassic Kayenta Formation of Arizona. It is anatomically very similar to modern frogs. The oldest known caecilian is another Early Jurassic species, *Eocaecilia micropodia*, also from Arizona. The earliest salamander is *Beiyuanerpeton jianpingensis* from the Late Jurassic of northeastern China [5].

Authorities disagree as to whether Salientia is a superorder that includes the order Anura, or whether Anura is a sub-order of the order Salientia. The Lissamphibia are traditionally divided into three orders, but an extinct salamander-like family, the Albanerpetontidae, is now considered part of Lissamphibia alongside the superorder Salientia. Furthermore, Salientia includes all three recent orders plus the Triassic proto-frog, *Triadobatrachus* [4,5].

An Artificial Neural Network (ANN) is a mathematical model that is driven by the functional feature of biological neural networks. A neural network contains an interconnected set of artificial neurons, and it processes information using a connectionist form to computation. As a rule, an ANN is an adaptive system that adjusts its structure based on external or internal data that runs over the network during the learning process. Current neural networks are non-linear numerical data modeling tools. They are usually used to model tricky relationships among inputs and outputs or to uncover patterns in data. ANN has been applied in several applications with significant accomplishment. For example, ANN has been effectively applied in the area of prediction, and handwritten character recognition [6].

Neurons are often come together into layers. Layers are groups of neurons that perform similar functions. There are three kinds of layers. The input layer is the layer of neurons that take input from the user program. The layer of neurons that send data to the user program is the output layer. Between the input layer and output layer there are hidden layers. Hidden layer neurons are connected only to other neurons and never directly interact with the user program. The input and output layers are not just there as interface points. Every neuron in a neural network has the opportunity to affect processing. Processing can occur at any layer in the neural network. Not every neural network has this many layers. The hidden layer is optional. The input and output layers are required, but it is possible to have a layer that act as both an input and output layer [7].

ANN learning can be either supervised or unsupervised. Supervised training is accomplished by giving the neural network a set of sample data along with the expected outputs from each of these samples. Supervised training is the most common form of neural network training. As supervised training proceeds, the neural network is taken through several iterations, or epochs, until the actual output of the neural network matches the expected output, with a reasonably small error. Each epoch is one pass through the training samples. Unsupervised training is similar to supervised training except that no expected outputs are provided. Unsupervised training usually occurs when the neural network is to classify the inputs into several groups[8].

The training progresses through many epochs, just as in supervised training. As training progresses, the classification groups are "discovered" by the neural network [9]. Training is the process by which these connection weights are assigned. Most training algorithms begin by assigning random numbers to the weight matrix. Then the validity of the neural network is tested. Next, the weights are adjusted based on validation results. This process is repeated until the validation error is within an acceptable limit [9]. Validation of the system is done once a neural network has been trained and it must be evaluated to see if it is ready for actual use. This final step is important so that it can be determined if additional training is required. To properly validate a neural network validation data must be set aside that is completely separate from the training data [10].

The Objective of this study is to propose an ANN model to predict the presence of amphibians species near the water reservoirs based on features obtained from GIS systems and satellite images.

2. Methodology

A data set refer to University of Silesia, Faculty of Earth Sciences, Sosnowiec 41-200, Poland [11] was used; it contains a number of factors that are considered to have an effect on the presence of amphibians species near the water. These factors were classified as input variables. The output variable represents the predicted amphibians based on those inputs.

2.1 The Input Variables

This database includes 189 cases. Each case represents the presence of amphibian's species. It was found that each of these cases belonged to one of seven classes.

Table 1: Input attributes

#	Attribute Name	Attribute Domain	Attribute Type
1	ID	Integer	Input
2	MV	Categorical	Input
3	SR	Numerical	Input
4	NR	Numerical	Input
5	TR	Categorical	Input
6	VR	Categorical	Input
7	SUR1	Categorical	Input
8	SUR2	Categorical	Input
9	SUR3	Categorical	Input
10	UR	Categorical	Input
11	FR	Categorical	Input
12	OR	Numerical	Input
13	RR	Ordinal	Input
14	BR	Ordinal	Input
15	MR	Categorical	Input
16	CR	Categorical	Input
17	Green frogs	Categorical; Label 1	Output
18	Brown frogs	Categorical; Label 2	Output
19	Common toad	Categorical; Label 3	Output
20	Fire-bellied toad	Categorical; Label 4	Output
21	Tree frog	Categorical; Label 5	Output
22	Common newt	Categorical; Label 6	Output
23	Great crested newt	Categorical; Label 7	Output

This variable, named Label 1 to label 7, represents the output variable. Except the output variable, there were 16 input variables for each amphibians species. Information of input and output variables are found in Table2.

Table2: Input attributes and its description

#	Attribute Name	Description of the attributes
1	ID	vector ID
2	MV	motorway
3	SR	Surface of water reservoir numeric [m2]
4	NR	Number of water reservoirs in habitat - Comment: The larger the number of reservoirs, the more likely it is that some of them will be suitable for amphibian breeding.
5	TR	Type of water reservoirs: <ul style="list-style-type: none"> a. reservoirs with natural features that are natural or anthropogenic water reservoirs (e.g., subsidence post-exploited water reservoirs), not subjected to naturalization b. recently formed reservoirs, not subjected to naturalization

		<ul style="list-style-type: none"> c. settling ponds d. water reservoirs located near houses e. technological water reservoirs f. water reservoirs in allotment gardens g. trenches h. wet meadows, flood plains, marshes i. river valleys j. streams and very small watercourses
6	VR	<p>Presence of vegetation within the reservoirs:</p> <ul style="list-style-type: none"> a. no vegetation b. narrow patches at the edges c. areas heavily overgrown d. lush vegetation within the reservoir with some part devoid of vegetation e. reservoirs completely overgrown with a disappearing water table <p>Comment: The vegetation in the reservoir favors amphibians, facilitates breeding, and allows the larvae to feed and give shelter. However, excess vegetation can lead to the overgrowth of the pond and water shortages.</p>
7	SUR1	Surroundings 1 "the dominant types of land cover surrounding the water reservoir
8	SUR2	Surroundings 2 "the second most dominant types of land cover surrounding the water reservoir
9	SUR3	<p>Surroundings 3 "the third most dominant types of land cover surrounding the water reservoir</p> <p>Comment: The surroundings feature was designated in three stages. First, the dominant surroundings were selected. Then, two secondary types were chosen.</p> <ul style="list-style-type: none"> a. forest areas (with meadows) and densely wooded areas b. areas of wasteland and meadows c. allotment gardens d. parks and green areas e. dense building development, industrial areas f. dispersed habitation, orchards, gardens g. river valleys h. roads, streets i. agricultural land <p>The most valuable surroundings of water reservoirs for amphibians are areas with the least anthropopressure and proper moisture.</p>
10	UR	<p>Use of water reservoirs:</p> <ul style="list-style-type: none"> a. unused by man (very attractive for amphibians) b. recreational and scenic (care work is performed) c. used economically (often fish farming) d. technological
11	FR	<p>The presence of fishing:</p> <ul style="list-style-type: none"> a. lack of or occasional fishing b. intense fishing c. breeding reservoirs <p>Comment: The presence of a large amount of fishing, in particular predatory and intense fishing, is not conducive to the presence of amphibians.</p>
12	OR	<p>Percentage access from the edges of the reservoir to undeveloped areas (the proposed percentage ranges are a numerical reflection of the phrases: lack of access, low access, medium access, large access to free space):</p> <ul style="list-style-type: none"> a. 0 < 25% lack of access or poor access b. 25 < 50% low access c. 50 < 75% medium access, d. 75 < 100% large accesses to terrestrial habitats of the shoreline is in

		contact with the terrestrial habitat of amphibians.
13	RR	Minimum distance from the water reservoir to roads: a. <50 m b. 50 < 100 m c. 100 < 200 m d. 200 < 500 m e. 500 < 1000 m f. > 1000 m Comment: The greater the distance between the reservoir and the road, the more safety for amphibians.
14	BR	Building development - Minimum distance to buildings: a. <50 m b. 50 < 100 m c. 100 < 200 m d. 200 < 500 m e. 500 < 1000 m f. >1000 m Comment: The more distant the buildings, the more favorable the conditions for the occurrence of amphibians.
15	MR	Maintenance status of the reservoir: a. Clean b. slightly littered c. reservoirs heavily or very heavily littered Comment: Trash causes devastation of the reservoir ecosystem. Backfilling and leveling of water reservoirs with ground and debris should also be considered.
16	CR	Type of shore a. Natural b. Concrete Comment: A concrete shore of a reservoir is not attractive for amphibians. A vertical concrete shore is usually a barrier for amphibians when they try to leave the water.
17	Label 1	the presence of Green frogs
18	Label 2	the presence of Brown frogs
19	Label 3	the presence of Common toad
20	Label 4	the presence of Fire-bellied toad
21	Label 5	the presence of Tree frog
22	Label 6	the presence of Common newt
23	Label 7	the presence of Great crested newt

Handling non-numeric data, such as Boolean = {true, false}, is more difficult. However, nominal-valued variables can be represented numerically. Value **true** will be replaced with value 1, and value **false** will be replaced with value 0. We will not use variable id name in experiment, because this variable is unique for each case.

Once the most appropriate raw input data has been selected, it must be preprocessed; otherwise, the neural network will not produce accurate forecasts.

Transformation and normalization are two widely used preprocessing methods. Transformation involves manipulating raw data inputs to create a single input to a network, while normalization is a transformation performed on a single data input to distribute the data evenly and scale it into an acceptable range for the network. Acceptable range of value, in *JNN tool*, varies in the range between zero and one.

2.2 Data Normalization

Linear scaling of data is one of the methods of data normalization. Linear scaling requires that a minimum and maximum values associated with the facts for a single data input be found. Let's call these values X_{min} and X_{max} , respectively. The formula for transforming each data value to an input value X is:

$$X_i = (X_i - X_{min}) / (X_{max} - X_{min}) \quad \text{Eq.(1)}$$

Our desired range varies in the interval between zero and one. In this study we normalized all numeric variables to be in the range between 0 and 1. However, the output variable consists of seven values (labels) in binary system where each position the in seven digits represent the presence of a specific amphibians species or not. We converted the seven binary digits to one numeric value then the numeric value was normalized to be between zero and one. We used the equation eq. (1) for the normalization.

2.3 Building the ANN Model

We have used Just Neural Network (JNN) tool [57] to build a multilayer ANN model. The proposed model consists of 3 Layers: Input Layer with 16 nodes, one Hidden Layer with 4 nodes and Output Layer with one node as can be seen in Figure 3.

We have sat the parameters of the proposed model as follows: Learning Rate 0.6 and the Momentum to be 0.8, and Average Error rate to be 0.01 (as shown in Figure 2).

2.4 Evaluating the ANN model

The presence of amphibians species dataset consists of 189 samples with 23 attributes as in Table 1 and Table 2. We imported the preprocessed CSV file of the presence of amphibians species dataset into the JNN environment (as seen in Figure 1). We divided the imported dataset into two sets (Training and Validation) randomly using the JNN tool. The Training consists of approximately 67% (127 samples) and the validation set consists of 33% of the dataset (62 samples). After making sure that the parameter control was sat properly, we started training the ANN model and kept eye on the learning curve, loss error and validation accuracy. We trained the ANN model for 301 cycles. The best accuracy we got was 100% (as seen in Figure 4). We determined the most influential factors in the presence of amphibians species dataset as in Figure 5. Figure 6 shows the summary of the proposed model.

Figure 1: Imported dataset into JNN environment

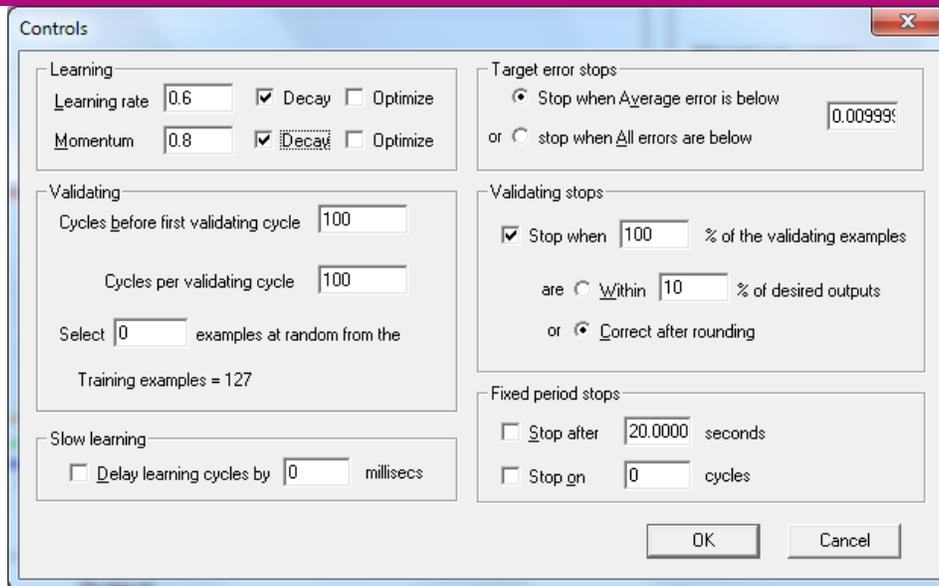


Figure 2: Control of the parameters of the proposed ANN model

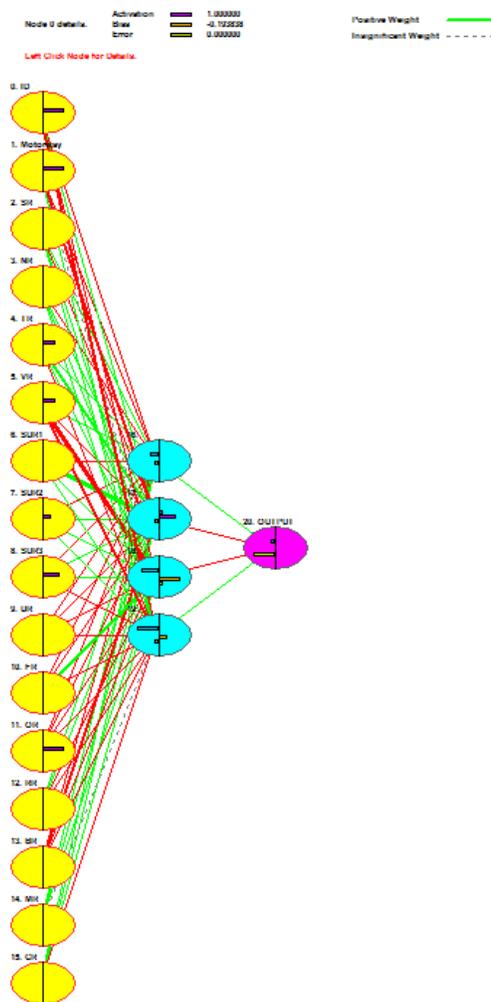


Figure 3: architecture of the final ANN model

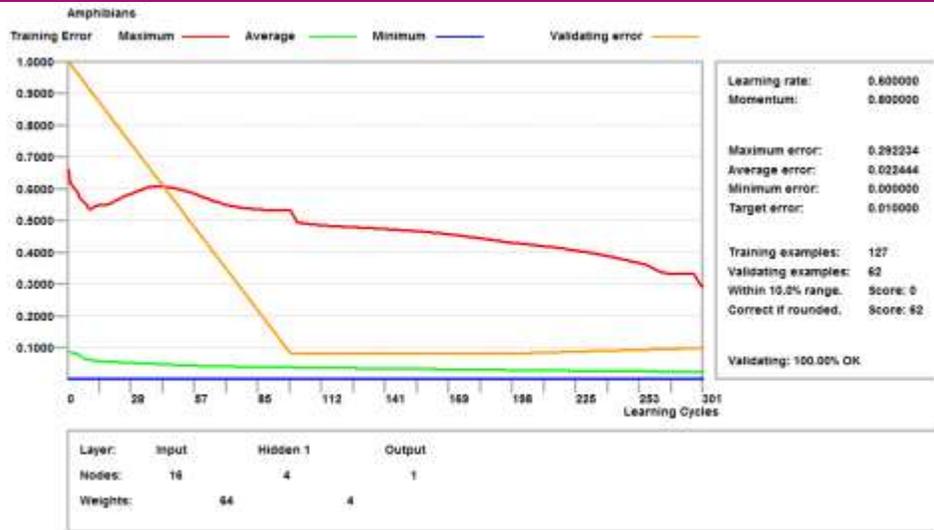


Figure 4: Training and validation curves of the proposed ANN model

Amphibians 301 cycles. Target error 0.0100 Average training error 0.022444

The first 16 of 16 Inputs in descending order.



Figure 5: The most influential Feature in the proposed ANN model

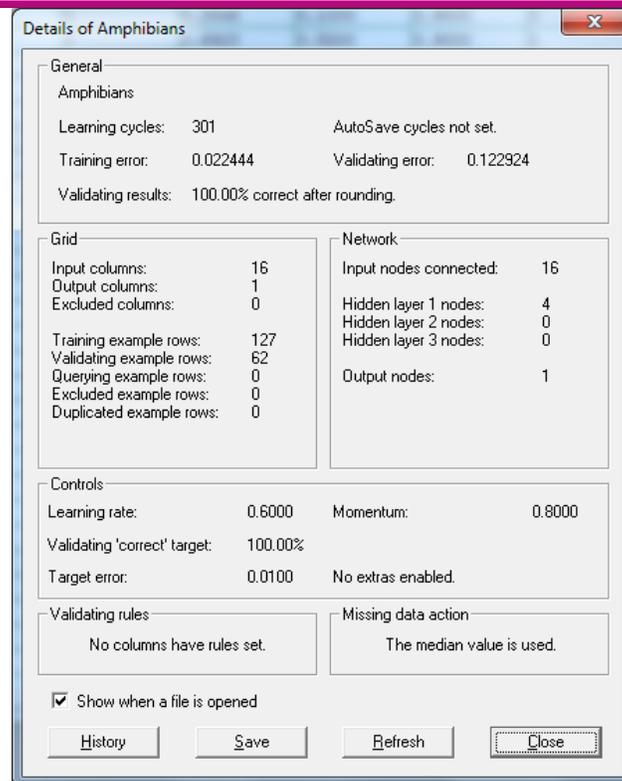


Figure 6: Details of the proposed ANN model

3. Conclusion

An Artificial Neural Network model for predicating presence of amphibian species prediction using features obtained from GIS and satellite images was presented. The model used feed forward backpropagation algorithm for training the proposed ANN model using JNN tool. The factors for the model were obtained from dataset which represents amphibian features of each amphibian species. The model was tested and the accuracy rate was 100%. This study showed that artificial neural network is capable of predicating amphibian species accurately.

References

1. Blackburn, D. C.; Wake, D. B. (2011). "Class Amphibia Gray, 1825. In: Zhang, Z.-Q. (Ed.) Animal biodiversity: An outline of higher-level classification and survey of taxonomic richness". Zootaxa. 3148: 39–55.
2. Baird, Donald (May 1965). "Paleozoic lepospondyl amphibians". Integrative and Comparative Biology. 5 (2): 287–294.
3. Frost, Darrel (2013). "American Museum of Natural History: Amphibian Species of the World 5.6, an Online Reference". The American Museum of Natural History.
4. <http://amphibiaweb.org:8000/lists/index.shtml>
5. Speer, B. W.; Waggoner, Ben (1995). "Amphibia: Systematics". University of California Museum of Paleontology.
6. White., F. The AETFAT chorological classification of Africa: history, methods and applications. Bull. Jard. Bot. Belg. 225–281 (1993).
7. Morrone, J. J. Biogeographical regionalisation of the Neotropical region. Zootaxa 3782, 1–110 (2014).
8. Hughes, C. E., Pennington, R. T. & Antonelli, A. Neotropical plant evolution: assembling the big picture. Bot. J. Linn. Soc. 171, 1–18 (2013).
9. Kappen., W. Versuch einer klassifikation der kli-mate, vorzugsweise nach ihrenbeziehungen zur pflanzen-welt. Geogr. Z. 6, 593–611 (1900).
10. Kottek, M., Grieser, J., Beck, C., Rudolf, B. & Rubel, F. World map of the Köppen-Geiger climate classification updated. Meteorol. Z. 15, 259–264 (2006).
11. Hagmeier, E. M. & Stults., C. D. A numerical analysis of the distributional patterns of north american mammals. Syst. Zool. 13, 125–155 (1964).
12. Kreft, H. & Jetz., W. A framework for delineating bio geographical regions based on species distributions. J. Biogeogr. 37, 2029–2053 (2010).

13. Holt, B. G. et al. An update of Wallace's zoogeographic regions of the world. *Science* 339, 74–78 (2013).
14. Koleff, P., Gaston, K. J. & Lennon, J. J. Measuring beta diversity for presence-absence data. *J. Anim. Ecol.* 72, 367–382 (2003).
15. Jaccard, P. Etude comparative de la distribution florale dans une portion des Alpes et du Jura. Impr. Corbaz (1901).
16. Dice, L. R. Measures of the amount of ecologic association between species. *Ecology* 26, 297–302 (1945).
17. Simpson, G. G. Mammals and the nature of continents. *Am. J. Sci.* 241, 1–31 (1943).
18. <https://archive.ics.uci.edu/ml/index.php>.
19. EasyNN Tool