World Journal of Biology Pharmacy and Health Sciences

eISSN: 2582-5542

Cross Ref DOI: 10.30574/wjbphs

Journal homepage: https://wjbphs.com/

(Research Article)

Check for updates

Amphibian occurrence prediction around water reservoirs: A machine learning perspective

Montaha A. A. Alammar *

Public Authority for Applied Education and Training (PAAET), Collage of Health science (CHS), Kuwait.

World Journal of Biology Pharmacy and Health Sciences, 2025, 22(01), 227-233

Publication history: Received on 26 February 2025; revised on 06 April 2025; accepted on 08 April 2025

Article DOI: https://doi.org/10.30574/wjbphs.2025.22.1.0388

Abstract

Amphibians play a crucial role in ecosystem stability; their presence or absence can serve as an indicator of environmental changes. This study investigates the application of machine learning (ML) models to predict amphibian species' occurrence, specifically the green frog, near water reservoirs. Data from satellite imagery and geographic information system (GIS) were analyzed to develop predictive models. Two classification models, Radial Basis Function (RBF) and Neural Network (NN)were evaluated alongside combination methods, including Bagging, Random Subspace Method (RSM), Boosting, and Feature-Based Combiner (FBC). Results revealed that the Neural Network combined with Boosting achieved the highest accuracy, with a classification rate nearly 70%, outperforming the RBF classifier in all combinations. The results highlight the effectiveness of combining classifiers techniques to enhance prediction accuracy and stability. These findings provide valuable insights into the potential of machine learning techniques for amphibian monitoring and ecological assessment.

Keywords: Amphibians; Species Prediction; Water Reservoirs; And Machine Learning

1. Introduction

Amphibians represent one of the most diverse groups of vertebrates, with abundant species populations worldwide [33]. They are four-limbed organisms belonging to an ancient lineage of vertebrates [30]. The term amphibian originates from the Greek word meaning "double life," referring to their ability to inhabit both aquatic and terrestrial environments throughout their life cycle [3]. However, water plays a central role in their biology, as they are primarily found near aquatic habitats [33]. Amphibians require high-quality water and suitable habitat conditions for survival, as they reproduce and develop in aquatic environments before transitioning to terrestrial habitats upon reaching maturity [6].

Amphibians are classified into three modern taxonomic orders: (I) Anura (frogs and toads), (II) Urodela (salamanders and newts), and (III) Apoda (caecilians) as shown in Figure 1 [13]. The largest order, Anura, includes over 7,000 species of frogs and toads, characterized by smooth, moist skin, long hind limbs, and large, protruding eyes. Their life cycle typically begins with eggs laid in water, which hatch into aquatic larvae (tadpoles) before undergoing metamorphosis into terrestrial adults. The Urodela order includes over 600 species of salamanders, which possess elongated, slender bodies, smooth skin, and four limbs. Like anurans, most salamanders reproduce in water, where their larvae undergo metamorphosis into adults. Some species, such as newts, exhibit both aquatic and terrestrial life stages. The Apoda order consists of more than 200 species of Caecilians, limbless amphibians with elongated, worm like bodies and smooth, slimy skin. They are primarily fossorial, living underground or in burrows, and many species give birth to live young that bypass metamorphosis. All amphibians rely primarily on cutaneous respiration, using their moist skin as a key respiratory organ for gas exchange [30]. This physiological trait also makes them highly sensitive to environmental

* Corresponding author: Montaha A. A. Alammar.

Copyright © 2025 Author(s) retain the copyright of this article. This article is published under the terms of the Creative Commons Attribution Liscense 4.0.

changes, including pollution and habitat degradation, making them important bioindicators of ecosystem health [28, 33].

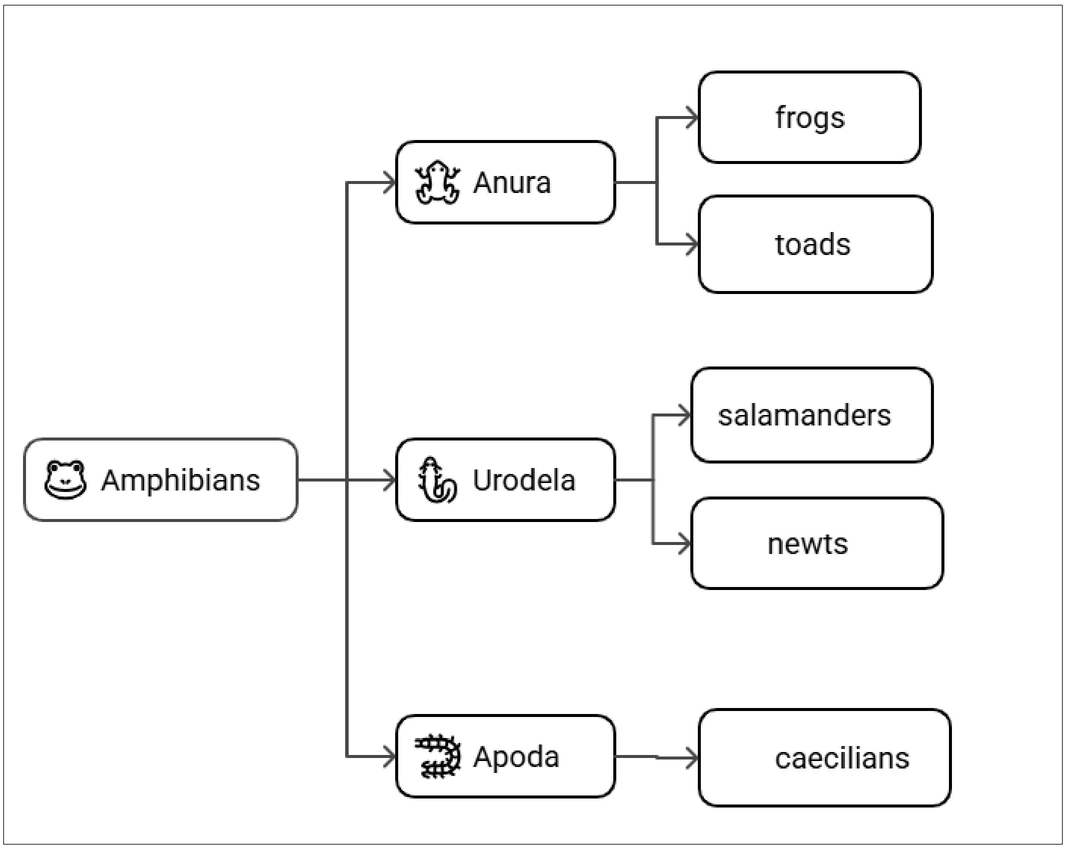

Figure 1 The Three Taxonomic orders of Amphibians

1.1. Water reservoirs

Water reservoirs play a key role in the natural environment that will affect the aquatic ecosystems [12]. These types of aquatic environment form an excellent home for various valuable ecological species of flora and fauna [31]. As mentioned earlier, amphibians have a preference to be closely located near water reservoirs as these habitats allow amphibians reproduction and act as a breeding site [2]. For example, some of the species (green frog) tend to stay in water reservoirs until autumn, given that their seasons last from the end of April to the end of June [5]. Additionally, Water reservoirs enhance the balance of species' thermal and hydric needs [25]. On the other hand, water reservoirs could help in field research as they are significant to accelerate the assessment of amphibian populations since it is easy to estimate the species number while they are in the water reservoirs [5].

1.2. Why is it important to predict the presence of amphibians?

Amphibians are generally approved to be valuable biological indicator species for assessing a variety of ecosystems throughout the world [32]. In ecological populations, amphibians can reveal both direct and indirect environmental impacts [6]. Hence, their implication in maintaining ecosystem stability makes them ideal species to function as biological indicators [27]. Their sensitivity of environmental changes improves understanding of both ecological and anthropogenic changes to the natural world [22]. Subsequently, since amphibians interact with their surrounding environment via their skin, they absorb water and other environmental compounds such as gases throughout their skin [27]. Hence, their moist, permeable skin makes them sensitive and exposed to toxic substances (including pesticides and other air pollutants) which reflect the health of a particular ecosystem [35].

The presence of amphibians is mostly indicated by the condition of the habitat itself [22]. As a result, their existence and/or absence implies that there are major factors affecting the normal conditions of the environment [27]. It is important at this point to summarize some of the features of amphibians that make their existence significant for studying the environmental challenges. The main features are : 1) Amphibians are fundamental components of aquatic and terrestrial food webs, serving as prey and predators[34]. 2) They are a critical part of the food chain and act as essential elements of ecosystems serving as both predators and prey [3]. 3) Many other animals' survival is affected by them [26]. 4) Providing food resources for a variety of animals from snakes to raptors [35].

1.3. Machine Learning

Machine learning (ML) has been widely used in various fields to predict outcomes or classify samples. ML models, including pattern recognition classifiers [36, 19, 23], can identify trends and relationships within data sets. These models enable the development of automated systems or software applications that can perform repetitive tasks, often more efficiently than humans. Therefore, the goal of this study is to experimentally investigate several methods for predicting the presence of amphibians. The performance of different systems and methods will be compared by calculating the classification rate, averaged over ten cross-validation trials.

The following sections describe the research method, followed by the experimental results and discussion. The paper concludes in Section 4.

2. Materials and methods

This section describes the machine learning (ML) methods and dataset used in this experiment. The experiment was simulated using MATLAB on a personal computer. The data was randomly partitioned into a training set and a test set using the 10-fold cross-validation method. Results were averaged to obtain the classification rate. The classification rate is calculated as the number of correctly classified samples divided by the total number of test samples.

2.1. Data set

In this study, repurposed data from the UCI Machine Learning Repository [9] was used to evaluate the methodology. The data was collected in Poland for environmental impact assessment (EIA) reports related to two planned road projects. These reports were primarily used to gather information on the amphibian population size at each of the 189 occurrence sites. The objective was to predict the presence of amphibian species near water reservoirs based on features derived from GIS systems and satellite images [16]. The dataset contains 14 features and 7 binary labels, each corresponding to a different species of frog. For this preliminary study, the experiment focused on detecting only the green frog. Therefore, we selected the label for the green frog and discarded the other six labels, making the dataset a single-class classification problem. If the green frog was present, a label of "1" was assigned; otherwise, a label of "0" was used. Additionally, the missing data was checked and found to be absent. Finally, the dataset was normalized to a range between 0 and 1.

2.2. Classifiers

The experiment was repeated using two commonly used classifiers: the Radial Basis Function (RBF) classifier and the Neural Network (NN) classifier [4]. For the radial basis classifier, the spread was set to 1. The neural network classifier used was a backpropagation network with the Levenberg-Marquardt training function. It consisted of three layers: the number of neurons in the first layer matched the number of features, the hidden (second) layer contained 5 neurons, and the output layer had neurons equal to the number of classes. The network was trained for 100 epochs, with a training error goal of 0.001.

2.3. Combiners

To improve classifier performance, a combination of classifiers was used. Four prominent classifier combiner methods were tested: Bagging [7], Random Subspace Method (RSM) (Ho, 1998), ARCX4 boosting [8], and a feature-based combiner (FBC).

Bagging, proposed by Breiman, generates multiple versions of a predictor or classifier using bootstrapping, and aggregates these versions to form a final classifier. In this study, 25 versions of the classifier were created, as recommended by Breiman. The total number of samples in each bootstrap set was the same as the original training set.

The RSM method aims to create diverse classifiers by assigning different features to each classifier. Each classifier is given a subset of features, randomly selected without replacement from the total feature set. This ensures that classifiers view the data space differently. In this experiment, 75% of the total available features were used for each classifier, with a total of 25 classifiers combined.

Boosting, an effective combiner method proposed by Freund (1995) and improved by Freund & Schapire (1997), focuses on difficult samples [10, 11]. It creates multiple classifiers and assigns samples to these classifiers randomly, with each sample having a weight based on the probability of being selected for inclusion. Boosting iteratively builds a new classifier by adjusting the weight of each sample based on the performance of the last built classifier. Misclassified samples have their weights increased, while the weights of correctly classified samples are decreased. Below is the procedure for a specific boosting method, ARCx4, proposed by Breiman (1998):

- Step 1: Initialize the weights of the samples to 1/N, where N is the number of samples in the training set.
- Step 2: Create the first classifier using a training subset created from the original training set by randomly selecting samples given the initial weights of samples.
- Step 3: Check the performance of the classifier to find misclassified samples.
- Step 4: For each sample: we count the number of classifiers that misclassified the sample, then divide by number of classifiers in the combiner to get a value p.
- Step 5: Calculate h=1+ p4 for each sample.
- Step 6: Find the new weights for each sample as h of a sample divided by sum of all h for all samples.
- Step 7: If maximum number of classifiers of a combiner is not reached create a new classifier using the new weights. Otherwise stop.
- Step 8: Repeat steps 3 to 7.

3. Results and Discussion

The experimental results provide valuable insights into the performance of different classifiers and combiner methods for amphibian species detection. As shown in Table 1 and Figure 2, Neural Network-based approaches consistently outperformed Radial Basis Function (RBF) classifiers across all combiner strategies. The highest classification rate was achieved by the Neural Network combined with the Boosting combiner (69.47%), followed closely by the Neural Network with RSM (68.95%), suggesting that these combinations are particularly effective for identifying green frog presence. A notable performance gap emerged between the two classifier types when paired with RSM and FBC combiners. While Neural Networks maintained relatively stable performance across all combiners (ranging from 68.42% to 69.47%) with a standard deviation of only 0.5%, RBF classifiers exhibited significant performance degradation, particularly with RSM (57.89%) and FBC (56.32%). These values represent a drop of over 10 percentage points compared to their Bagging and Boosting implementations. These findings highlight the superior consistency of Neural Networks and the limitations of RBF classifiers in this classification task.

Table 1 average classification rate and standard deviation of neural network and RBFN classifiers using four types of combiners.

Classifier	Combiner	Average Rate	Standard Dev.
Neural Network	Bagging	68.42%	6.56%
Neural Network	Boosting	69.47%	8.88%
Neural Network	RSM	68.95%	8.75%
Neural Network	FBC	68.42%	7.44%
Radial Basis Function	Bagging	66.32%	6.66%
Radial Basis Function	Boosting	65.79%	8.68%
Radial Basis Function	RSM	57.89%	4.96%
Radial Basis Function	FBC	56.32%	11.91%

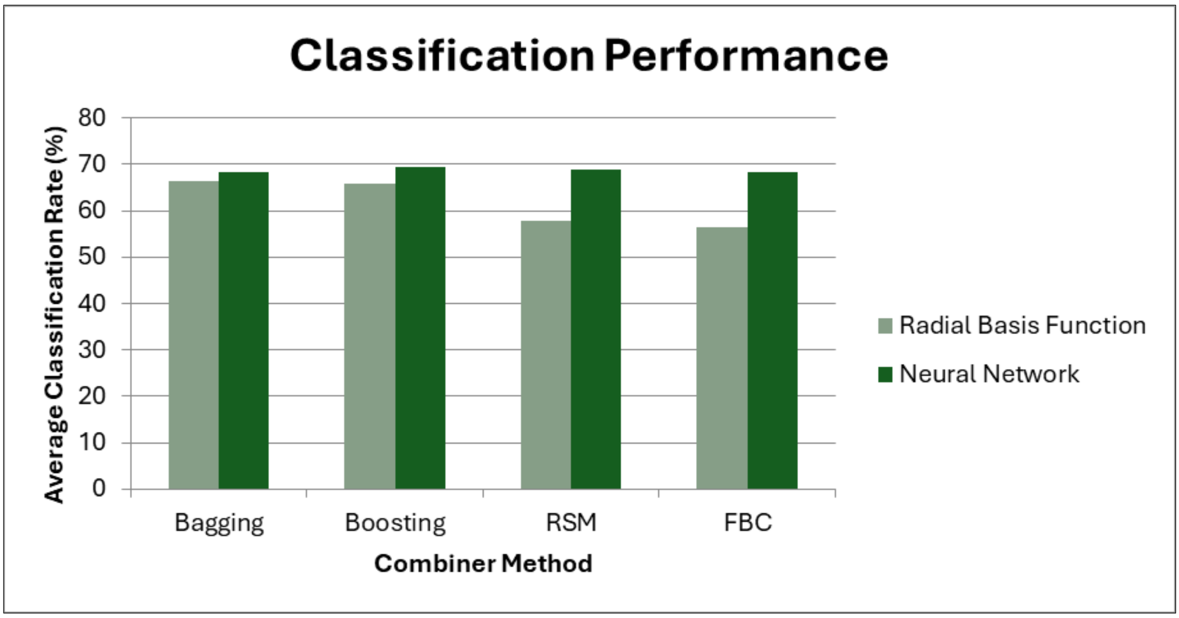

Figure 2 The chart presents the average classification rates (%) for Neural Network (NN) and Radial Basis Function (RBF) classifiers combined with Bagging, Boosting, RSM, and FBC

The standard deviation (SD) analysis provides additional insights into model stability. A low SD indicates consistent performance, while a high SD suggests greater variability. The RBF with FBC combination, while showing the lowest classification accuracy, exhibited the highest variability (SD = 11.91), indicating inconsistent predictions across the 10-fold cross-validation. Conversely, RBF with RSM showed the lowest standard deviation (4.96), suggesting that while its performance was poor, it was consistently so. This trade-off between performance and consistency appears less pronounced in Neural Network models, which maintained balanced standard deviations across combiner methods. Among the combiner methods, Boosting and Bagging produced the most favourable results overall, suggesting that these ensemble techniques are more effective for the classification task. However, the relatively high standard deviation observed with some combinations—particularly with the FBC combiner at 8.56%—highlights the challenge of achieving performance consistency, a concern that should be addressed in future work.

These findings suggest that for amphibian detection applications, Neural Network classifiers offer superior and more stable performance. Boosting is a strong choice for this task, providing marginal improvements over other combiner methods. Future studies could explore dimensionality reduction using feature selection techniques to improve class separation and enhance performance. Additionally, the RBF classifier may achieve better results when combined with Bagging. Further experiments with different hyperparameters and feature selection methods could lead to even greater classification accuracy.

4. Conclusion

In this study, we explored various machine learning methods for amphibian species detection near water reservoirs, utilizing backpropagation neural networks and radial basis function classifiers for sample classification. To enhance performance, we incorporated combiner methods such as Boosting and Bagging. Our results demonstrated that the Neural Network combined with Boosting achieved the highest performance, with an accuracy of nearly 70%. In contrast, the Radial Basis Function classifier with Bagging exhibited the lowest performance, underscoring the need for further optimization. The Boosting method consistently improved classification accuracy, reinforcing its robustness as a preferred technique for this task. Additionally, the relationship between average accuracy and standard deviation highlighted the importance of balancing performance with variability. While some classifiers achieved high accuracy with moderate variability, others maintained stable performance at lower accuracy levels. These findings offer valuable insights into classifier selection for future modelling efforts. Further experiments, particularly targeting different stages of the system, are recommended to optimize classification accuracy and enhance overall performance.

Compliance with ethical standards

Acknowledgments

The author would like to thank Dr. Fuad M. Alkoot – HITN - PAAET, for his assistance in setting up the program used for the experiments.

Disclosure of conflict of interest

No conflict of interest to be disclosed.

References

[1] Alkoot FM, Alqallaf AK. Investigating machine learning techniques for the detection of autism. International Journal of Data Mining and Bioinformatics. 2016;16(2):141-69.

[2] Baker J, Beebee T, Buckley J, Gent T, Orchard D. Amphibian habitat management handbook. Amphibian and reptile conservation, Bournemouth. 2011;39.

[3] Bartelt PE, Thornton PE, Klaver RW. Modelling physiological costs to assess impacts of climate change on amphibians in Yellowstone National Park, USA. Ecological Indicators. 2022 Feb 1;135:108575.

[4] Bishop CM, Nasrabadi NM. Pattern recognition and machine learning. New York: springer; 2006 Aug 17.

[5] Blachnik M, Sołtysiak M, Dąbrowska D. Predicting presence of amphibian species using features obtained from gis and satellite images. ISPRS International Journal of Geo-Information. 2019 Mar 1;8(3):123.

[6] Blaustein AR, Walls SC, Bancroft BA, Lawler JJ, Searle CL, Gervasi SS. Direct and indirect effects of climate change on amphibian populations. Diversity. 2010 Feb 25;2(2):281-313.

[7] Breiman L. Bagging predictors. Machine learning. 1996 Aug; 24:123-40.

[8] Breiman L. Arcing classifier (with discussion and a rejoinder by the author). The annals of statistics. 1998 Jun;26(3):801-49.

[9] Dua D, Graff C. UCI machine learning repository [http://archive.ics.uci.edu/ml]. Irvine, CA: University of California, School of Information and Computer Science. IEEE transactions on pattern analysis and machine intelligence. 2019;1(1):1-29.

[10] Freund Y, Schapire RE. A decision-theoretic generalization of on-line learning and an application to boosting. Journal of computer and system sciences. 1997 Aug 1;55(1):119-39.

[11] Freund Y, Schapire R, Abe N. A short introduction to boosting. Journal-Japanese Society For Artificial Intelligence. 1999 Sep 1;14(771-780):1612.

[12] Gardner TA, Barlow J, Peres CA. Paradox, presumption and pitfalls in conservation biology: the importance of habitat change for amphibians and reptiles. Biological conservation. 2007 Aug 1;138(1-2):166-79.

[13] Guala G. Integrated Taxonomic Information System (ITIS).

[14] Gutiérrez DR. Introductory Chapter: General Features of Reptiles and Amphibians. InReptiles and Amphibians 2018 Jul 18. IntechOpen.

[15] Blachnik M, Sołtysiak M, Dąbrowska D. Predicting presence of amphibian species using features obtained from gis and satellite images. ISPRS International Journal of Geo-Information. 2019 Mar 1;8(3):123.

[16] Habib NS, Abu Maghasib OK, Al-Ghazali AR. Predicting the Presence of Amphibian Species Using Features Attained from GIS and Satellite Images. 2021.

[17] Habib NS, Maghasib OK, Al-Ghazali AR, Abu-Nasser BS, Abu-Naser SS. Presence of Amphibian Species Prediction Using Features Obtained from GIS and Satellite Images. International Journal of Academic and Applied Research (IJAAR). 2020;4(11).

[18] Hamer AJ, McDonnell MJ. Amphibian ecology and conservation in the urbanising world: a review. Biological conservation. 2008 Oct 1;141(10):2432-49.

[19] Han R, Liu P, Wang G, Zhang H, Wu X. Advantage of Combining OBIA and Classifier Ensemble Method for Very High-Resolution Satellite Imagery Classification. Journal of Sensors. 2020;2020(1):8855509.

[20] Ho TK. The random subspace method for constructing decision forests. IEEE transactions on pattern analysis and machine intelligence. 1998 Aug 31;20(8):832-44.

[21] Hocking DJ, Babbitt KJ. Amphibian contributions to ecosystem services. Herpetological conservation and biology. 2014.

[22] Hopkins WA. Amphibians as models for studying environmental change. ILAR journal. 2007 Jan 1;48(3):270-7.

[23] Li F, Fan Y, Zhang X, Wang C, Hu F, Jia W, Hui H. Multi-feature fusion method based on EEG signal and its application in stroke classification. Journal of medical systems. 2020 Feb; 44:1-1.

[24] Muths E, Jung RE, Bailey LL, Adams MJ, Corn PS, Dodd Jr CK, Fellers GM, Sadinski WJ, Schwalbe CR, Walls SC, Fisher RN. Amphibian Research and Monitoring Initiative (ARMI): a successful start to a national program in the United States. USGS Staff-Published Research. 2005 Jan 1:129.

[25] Ortega Z, Ganci CC, Rivas ML. Thermoregulation and hydric balance in amphibians. InEvolutionary ecology of amphibians 2023 Jun 30 (pp. 103-119). CRC Press.

[26] Pierce Jr RJ. Lujan v. Defenders of Wildlife: Standing as a Judicially Imposed Limit on Legislative Power. Duke LJ. 1992; 42:1170.

[27] Rohman F, Permana H, Akhsani F, Wangkulangkul S, Priambodo B. The amphibian’s diversity as bioindicator of aquatic ecosystem at Sumber Taman, Malang, East Java. InAIP Conference Proceedings 2020 Apr 21 (Vol. 2231, No. 1). AIP Publishing.

[28] Shoemaker V, Nagy KA. Osmoregulation in amphibians and reptiles. Annual review of physiology. 1977 Mar;39(1):449-71.

[29] Halliday T, Adler K, editors. The new encyclopedia of reptiles and amphibians. Oxford, UK: Oxford University Press; 2002.

[30] Hibbitts TJ, LaDuc TJ, Hibbitts TL, Hibbitts TD. Texas amphibians: A field guide. University of Texas Press; 2012 Dec 31.

[31] Unglaub B, Steinfartz S, Kühne D, Haas A, Schmidt BR. The relationship between habitat suitability, population size and body condition in a pond-breeding amphibian. Basic and Applied Ecology. 2018 Mar 1; 27:20-9.

[32] Waddle JH. Use of amphibians as ecosystem indicator species (Doctoral dissertation, University of Florida).

[33] Wake DB, Koo MS. Amphibians. Current Biology. 2018 Nov 5;28(21): R1237-41.

[34] Wells KD. The ecology and behavior of amphibians. University of Chicago press; 2019 Dec 31.

[35] West J. Importance of amphibians: A synthesis of their environmental functions, benefits to humans, and need for conservation.

[36] Yaghoubi V, Cheng L, Van Paepegem W, Kersemans M. An ensemble classifier for vibration-based quality monitoring. Mechanical Systems and Signal Processing. 2022 Feb 15; 165:108341.