



From calls to conservation: AnuraPAM, a ready-to-use acoustic classifier for Central European anurans

BENJAMIN ELSNER^{1,*}, FLORIAN MEERHEIM^{1,2,*}, FRANK DZIOCK¹ & RAFFAEL ERNST^{3,4}

¹ HTWD University of Applied Sciences, Bioacoustics working group, Pillnitzer Platz 2, 01326 Dresden, Germany

² TUD Dresden University of Technology, Faculty of Environmental Sciences, 01062 Dresden, Germany

³ Senckenberg – Leibniz Institution for Biodiversity and Earth System Research, Königsbrücker Landstraße 159, 01109 Dresden, Germany

⁴ TUD Dresden University of Technology, Faculty of Biology, 01062 Dresden, Germany

* These authors contributed equally

Corresponding author: RAFFAEL ERNST, e-mail: raffael.ernst@senckenberg.de

Manuscript received: 9 June 2026

Accepted: 27 July 2026 by JÖRN KÖHLER

Abstract. Freshwater ecosystems are among the most threatened habitats worldwide, and amphibians represent the most endangered vertebrate group globally. Effective conservation strategies therefore require scalable, cost-efficient monitoring methods. Passive acoustic monitoring (PAM), combined with automated call classification, has emerged as a promising approach, yet amphibians remain underrepresented compared to birds. Here, we present AnuraPAM, a flexible, ready-to-use collection of automated call classifier models for the Central European anuran fauna, trained using the deep-learning framework BirdNET. We developed four convolutional neural network-based classification models for 14 Central European amphibian species including one hybrid taxon (*Pelophylax* kl. *esculentus*), using different training data strategies. Model performance was evaluated using one-vs-rest receiver operating characteristic (ROC) and precision–recall (PR) curves. Most species achieved high classification performance, with ROC-AUC (area under the curve) values exceeding 0.9 across model variants. However, classification accuracy varied among species, particularly within the *Pelophylax* complex, where hybridization poses substantial challenges. Increasing the amount of training data did not consistently improve model performance, highlighting the importance of data quality and representativeness. This is supported by the fact that our best-performing model performed better than two freely available classifiers, FrogNET and Perch v2, despite being trained on smaller datasets. Our results demonstrate that automated acoustic identification of Central European anurans is feasible and reliable for most species. The developed classifiers provide a novel tool for large-scale amphibian monitoring and offer considerable potential for application in conservation practice and citizen science projects. All our models are available as ready-to-use classifiers for the application in free audio analysis software such as BirdNET-Analyzer, PAMalyzer and Raven Intelligence.

Key words. Amphibia, Anura, bioacoustics, BirdNET, convolutional neural networks, passive acoustic monitoring, conservation, citizen science.

Introduction

Global biodiversity loss represents one of the most pressing environmental challenges of the 21st century. Current estimates suggest that approximately one million species are threatened with extinction, many within the coming decades (IPBES 2019). Extinction rates far exceed background levels, leading to the characterization of the present era as the sixth mass extinction (BARNOSKY et al. 2011, BANKS-LEITE et al. 2020).

Freshwater ecosystems are disproportionately affected by these trends (SAYER et al. 2025). Between 1970 and 2020, freshwater vertebrate populations declined by 85%, exceeding losses observed in terrestrial and marine sys-

tems (WWF 2024). Amphibians, which depend strongly on freshwater habitats, show the highest proportion of threatened species among vertebrates, with 41% assessed as threatened on the global IUCN Red List (IUCN 2026).

These global patterns are also evident in Europe, where nearly half of all freshwater habitat types are considered threatened (JANSSEN et al. 2016). Effective amphibian conservation therefore requires robust and scalable monitoring approaches, yet traditional survey methods are often time-consuming, costly, and spatially limited.

PAM has gained increasing attention as a non-invasive and cost-efficient alternative (GIBB et al. 2019). Autonomous recording units enable long-term data collection across large spatial scales, but the resulting datasets neces-

sitate automated analysis tools (SUGAI et al. 2019). Convolutional neural networks (CNNs) have proven particularly effective for bioacoustic classification (GIBB et al. 2019).

BirdNET is a CNN-based framework originally developed for bird call identification and now widely used in research and citizen science (KAHL et al. 2021). While numerous automated classifiers exist for birds (KNIGHT et al. 2017, STOWELL et al. 2019, COLE et al. 2022), amphibians remain underrepresented despite their high conservation relevance. So far, available automated classifiers for amphibians (HUANG et al. 2009, COLONNA et al. 2016, HASSAN et al. 2017, ANUNCIAÇÃO et al. 2022, DE ALMEIDA et al. 2024) did not cover Central Europe with two recent exceptions, namely FrogNET (KIRKLAND 2025) and Perch v2 (VAN MERRIËNBOER et al. 2026). These classifiers were trained on large uncurated data sets and their performance remains to be tested. Here, we present a customized, ready-to-use BirdNET-based classifier for 14 Central European anuran species including one hybrid taxon, based on a rare, well-curated call dataset compiled over several decades.

We specifically address two connected general research questions: (1) Is it possible to train a BirdNET-based model that reliably distinguishes 14 Central European amphibian species? (2) How does the amount of training data per species influence classification performance? We then compare the performance of our classifiers with the only two available classifiers that cover our focal taxa, FrogNET and Perch v2 and discuss the results with regard to the integration of the newly developed classifiers into standard monitoring routines for anurans and their suitability for citizen science projects. On this basis, we make recommendations for refinements and standardized tests under real-world conditions.

Material and methods

Study species and call data

The classifier was trained on call recordings of 14 amphibian species (see Table 1) native to Central Europe, including the hybrid taxon *Pelophylax* kl. *esculentus* (summarized in Table 1). The exotic American bullfrog (*Aquarana catesbeiana*), which must now be considered part of the Central European frog fauna (FICETOLA et al. 2007), was excluded due to an insufficient amount of available call data. *Pelophylax* kl. *esculentus* originates from hybridogenesis between *P. lessonae* and *P. ridibundus* and reproduces via hemiclinal inheritance (HOFFMANN & REYER 2013, KRIZMANIĆ & IVANOVIĆ 2010). Hybrid individuals are morphologically and acoustically intermediate and acoustic properties may be influenced by the respective local hybrid system, complicating acoustic identification (DE ALMEIDA et al. 2024). The distribution ranges of the species included in our analyses extend across Europe with significant interspecific range differences (wide-spread, e.g., *Bufo bufo* and *Rana temporaria* or more range-restricted, e.g., *Alytes obstetricans* and *Epidalea calamita*). When selecting call recordings for analyses, we specifically chose

Table 1. Overview of the anuran species included in AnuraPAM and their status in the Red List of Germany (Rote-Liste-Gremium Amphibien und Reptilien 2020) and Europe (TEMPLE et al. 2009); Nomenclature follows FROST (2026). Categories described in the German Red List: D = data insufficient, * = least concern, V = near threatened, G = threatened (unknown extent), 3 = vulnerable, 2 = endangered; category described in the European Red List: LC = least concern.

Taxon	Red List status Germany	Red List status Europe
<i>Alytes obstetricans</i> (LAURENTI, 1768)	2	LC
<i>Bombina bombina</i> (LINNAEUS, 1761)	2	LC
<i>Bombina variegata</i> (LINNAEUS, 1758)	2	LC
<i>Bufo bufo</i> (LINNAEUS, 1758)	*	LC
<i>Bufo viridis</i> (LAURENTI, 1768)	2	LC
<i>Epidalea calamita</i> (LAURENTI, 1768)	2	LC
<i>Hyla arborea</i> (LINNAEUS, 1758)	3	LC
<i>Pelobates fuscus</i> (LAURENTI, 1768)	3	LC
<i>Pelophylax</i> kl. <i>esculentus</i> (LINNAEUS, 1758)	*	LC
<i>Pelophylax lessonae</i> (CAMERANO, 1882)	G	LC
<i>Pelophylax ridibundus</i> (PALLAS, 1771)	D	LC
<i>Rana arvalis</i> NILSSON, 1842	3	LC
<i>Rana dalmatina</i> FITZINGER, 1838	V	LC
<i>Rana temporaria</i> LINNAEUS, 1758	V	LC

calls from populations of Central European origin in accordance with delimitations given in SILLERO et al. (2014) and COGĂLNICEANU et al. (2024). In addition to the focus taxa calls, three acoustically confusable non-amphibian species (*Gryllotalpa gryllotalpa*, *Caprimulgus europaeus*, *Otus scops*) were included in the call data set as “confusion classes”.

Recordings were obtained from three different sources: (1) private contributors (unpublished call collections: K. GROSSENBACHER, A. KWET, B. TRAPP); (2) published audio collections accompanying field guides (CREEMERS & VAN DELFT 2009, SCHNEIDER 2005, TETZLAFF 2007); and (3) online sound archives, namely Xeno-canto (Xeno-canto Foundation 2026) and the Sonothèque du Muséum national d'Histoire naturelle (Muséum national d'Histoire naturelle 2026). The call collection of K. GROSSENBACHER is unique because it represents the most extensively curated dataset incorporated in this study in terms of precise call metadata, including not only exact locality but also air and water temperature at time of recording. It comprises 78 independent recordings from 54 localities, covering all 14 focal taxa across 11 European countries and spanning a recording period of 52 years.

Call data preparation

All recordings were manually screened to confirm species identity and to ensure that a representative call repertoire

of each species is included. Recordings with excessive background noise or other major disturbances were excluded. If the files were not already in WAV format, they were converted and structured into species specific folders. Subsequently 25% of the recordings of each species including the confusion species were separated to form an independent test dataset. The remaining data were used for training. A selection of 10 representative sound examples including nine target species and one confusion species is available in Zenodo (<https://doi.org/10.5281/zenodo.20525666>). The majority of the data used are subject to specific copyright licenses; therefore, the full call dataset cannot be made freely available here.

The training files were manually screened and sequences deemed fitting for the training dataset were cut into three-second segments containing vocalizations of target or confusion species and gathered in species-specific folders. During the segmentation excerpts with dominant background noise, poor audio quality or overlapping anuran vocalizations of different species were excluded while typical background noise within the files was extracted to form a separate noise class for three models. For the fourth model, model “Raw”, additional soundscape files were acquired from Xeno-canto and left uncut to form the noise class. A minimum of at least 150 segments were prepared for each target species class. Additionally, two species-specific negative classes containing segments of confusion species for anurans, being *Otus scops* for *Alytes obstetricans* and *Caprimulgus europaeus* as well as *Gryllotalpa gryllotalpa* for *Bufotes viridis*, were created. The final training dataset for each model comprised 17 classes: 14 target species including one hybrid taxon, two negative classes and one noise class. The final number of three-second segments per species was determined during manual segmentation based on the judgement of the annotator, aiming to achieve a representative training dataset for each species. Consequently, the number of extracted segments varied between species and is not directly linked to the quantity of the available raw recordings. Table 2 provides a detailed overview of the number of cut segments per class and the number of source files from which they were extracted for each of our models, as well as for the test dataset. The separated test dataset contained a total of 199 recordings and was manually annotated using PAMalyzer, a custom version of AviaNZ and software for audio analysis in which the BirdNET-Analyzer is implemented (MARSLAND 2019, MEERHEIM et al. unpubl. data). We marked all identifiable sound events of target anuran species. Included recordings of the confusion species were not annotated to serve as distraction factors during evaluation.

AnuraPAM model training and evaluation

Four classification models were trained using the BirdNET Analyzer v2.4.0 (KAHL et al. 2021): Model 100 was trained using 100 randomly selected call segments per class and Model 150 used 150 randomly selected call segments per class. Model CompSeg was trained using all

Table 2. Overview of the quantity of three-second segments per class in each model, as well as the number of different files the segments were taken from.

Model	Three-second segments/files				
	100	150	CS	Raw	Test
<i>A. obstetricans</i>	100/23	150/25	283/26	734/31	262/13
<i>B. bombina</i>	100/19	150/21	211/21	613/28	375/11
<i>B. variegata</i>	100/27	150/27	234/27	1152/40	515/13
<i>B. bufo</i>	100/19	150/21	162/22	1445/25	179/13
<i>B. viridis</i>	100/27	150/32	185/34	733/43	366/14
<i>E. calamita</i>	100/19	150/19	217/19	1018/37	339/13
<i>H. arborea</i>	100/23	150/25	190/25	1296/58	584/29
<i>P. fuscus</i>	100/12	150/12	158/12	538/15	88/5
<i>P. kl. esculentus</i>	100/8	150/12	158/12	398/14	147/5
<i>P. lessonae</i>	100/15	150/15	158/15	780/35	596/17
<i>P. ridibundus</i>	100/25	150/25	189/27	2717/77	1221/28
<i>R. arvalis</i>	100/18	150/19	158/19	641/28	130/9
<i>R. dalmatina</i>	100/12	150/13	168/13	1249/25	307/8
<i>R. temporaria</i>	100/17	150/17	152/17	4371/34	488/12
<i>C. europaeus</i>	50/33	75/36	289/40	1268/40	–
<i>G. gryllotalpa</i>	50/28	75/31	211/32	1948/49	–
<i>O. scops</i>	100/22	150/22	186/22	436/22	–
Noise	100/27	150/35	150/35	1475/20	1216/199

manually prepared call segments per class and Model Raw was trained using all available raw unsegmented recordings per class.

We used the autotune option for hyperparameters. For all other training parameters default settings were retained, as they are appropriate for most training applications (SYMES et al. 2023). Only the crop mode for the model Raw was changed to “segments”, which changes the usage from a single segment to consecutive three-second segments per file for training. Subsequently each trained model was applied to classify the test dataset using PAMalyzer. Again, we used default settings for sensitivity (1.0) and overlap (0.0) and a minimum confidence threshold of 0.01 was chosen to keep all possible detections. Classification outputs, as well as the ground truth annotations were then exported as tabular data. Evaluation was approached using segment-based comparison of the ground truth annotations and predictions given by the models (VAN MERRIËNBOER et al. 2024). To assess model performance ROC curves and PR curves were calculated for each target class, as well as their respective AUC values. For this a one-vs-rest approach was applied treating each class as the positive class and all other classes as negatives. Macro-AUC values were calculated as arithmetic means across species.

Comparison of AnuraPAM with existing classifiers

We used the identical approach to test and evaluate the performance of FrogNET and Perch v2 in comparison to

our trained four AnuraPAM models. For the application of FrogNET we converted the available model from TensorFlow.js to TensorFlow Lite format and used it in PAMalyzer. Perch v2 was applied in BirdNET-Analyzer v2.4.0. We ensured that our test data is independent from our training data. Recordings in our test dataset were mainly obtained from Xeno-canto. As both Perch v2 and FrogNet are trained with recordings from Xeno-canto, it is highly likely that there is an overlap between the training data for these two classifiers and our test dataset. This overlap may represent a classification advantage for these models and could result in an overestimation of their performance compared to our models' performance. Hence AnuraPAM model performance rather represents a conservative estimate. However, it should be noted that both FrogNET and Perch v2 cover a wider range of anuran species and the performance metrics are calculated only on the 14 target species covered by the classifiers presented in this study. While FrogNET was trained as a custom classifier based

on BirdNET to differentiate between 21 European species, Perch v2 is a global classifier covering over 14000 species worldwide, including more than 1300 anurans. In contrast to these classifiers, we defined species specific negative classes with confusion species during training, which, in addition to data quality, may result in an even larger performance gap.

Results

ROC-AUC performance

All four AnuraPAM models achieved high overall performance (Fig. 1). Macro ROC-AUC values ranged from 0.926 (Model CompSeg) to 0.973 (Model Raw). Most species achieved ROC-AUC values above 0.910 across all models. Slightly lower values were only observed for *Bombina variegata* and within the *Pelophylax* complex (Table 3).

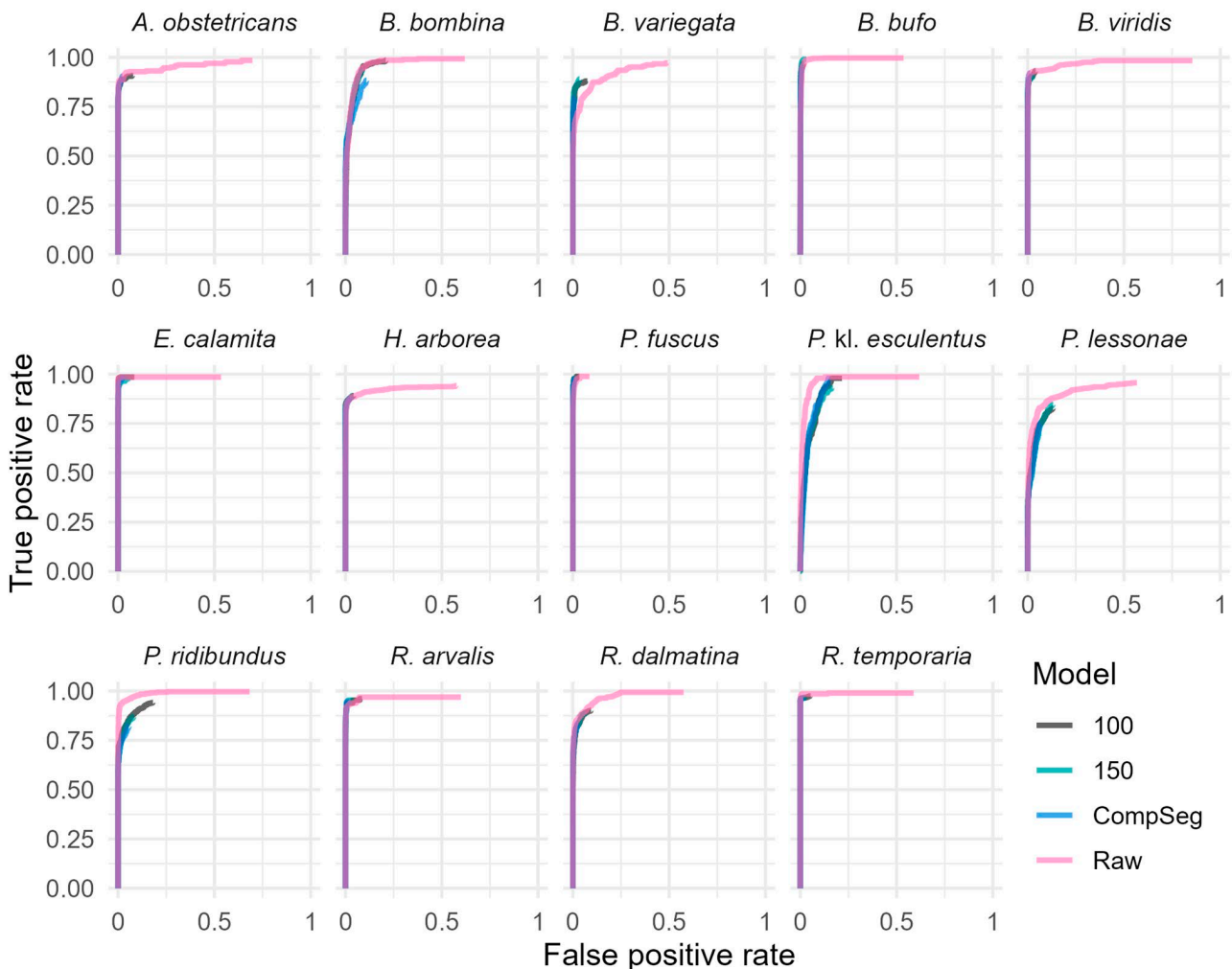


Figure 1. ROC curves per species of the four different AnuraPAM models evaluated on an independent test data set; the upper right point of each curve corresponds to the minimum confidence level of 0.01 provided by BirdNET.

Table 3. Calculated ROC-AUC and PR-AUC values per species for the AnuraPAM models trained (model 100, model 150, model CompSeg (CS) and model Raw); highest and lowest values per species are marked in bold and italic characters, respectively.

Model	ROC-AUC				PR-AUC			
	100	150	CS	Raw	100	150	CS	Raw
<i>Alytes obstetricans</i>	0,949	<i>0,934</i>	0,950	0,964	0,894	<i>0,881</i>	0,899	0,910
<i>Bombina bombina</i>	0,970	0,953	<i>0,921</i>	0,972	<i>0,705</i>	0,751	0,713	0,734
<i>Bombina variegata</i>	0,933	0,942	<i>0,897</i>	0,944	0,838	0,854	0,795	<i>0,724</i>
<i>Bufo bufo</i>	0,984	0,994	<i>0,974</i>	0,995	0,956	0,976	<i>0,943</i>	0,973
<i>Bufotes viridis</i>	0,963	<i>0,950</i>	0,960	0,974	0,924	<i>0,910</i>	0,929	0,932
<i>Epidalea calamita</i>	0,992	<i>0,988</i>	0,990	0,989	0,982	<i>0,969</i>	0,976	0,985
<i>Hyla arborea</i>	0,942	0,943	<i>0,939</i>	0,942	0,907	0,907	<i>0,901</i>	0,902
<i>Pelobates fuscus</i>	0,994	<i>0,988</i>	0,988	0,998	0,970	0,958	0,965	<i>0,936</i>
<i>Pelophylax kl. esculentus</i>	0,948	<i>0,927</i>	0,950	0,975	0,382	0,336	<i>0,325</i>	0,666
<i>Pelophylax lessonae</i>	0,886	0,892	<i>0,856</i>	0,931	0,707	0,696	<i>0,673</i>	0,790
<i>Pelophylax ridibundus</i>	0,953	0,924	<i>0,902</i>	0,991	0,915	0,888	<i>0,864</i>	0,978
<i>Rana arvalis</i>	0,978	0,976	<i>0,972</i>	0,973	0,930	0,934	0,934	<i>0,929</i>
<i>Rana dalmatina</i>	0,941	0,937	<i>0,915</i>	0,979	0,832	0,843	<i>0,828</i>	0,876
<i>Rana temporaria</i>	0,987	0,984	<i>0,982</i>	0,991	0,976	0,974	<i>0,971</i>	0,987
Mean	0,945	0,952	0,926	0,973	0,828	0,848	<i>0,807</i>	0,880

PR-AUC performance

PR-AUC values revealed stronger differences among species (Fig. 2). Again, model Raw achieved the highest macro-PR-AUC value (0.880) while the model trained with all available segments (Model CompSeg) showed the lowest PR-AUC value (0.807). Several species (*Bufo bufo*, *Epidalea calamita*, *Pelobates fuscus*, *Rana temporaria*) showed consistently high PR-AUC values, whereas *Pelophylax kl. esculentus* exhibited very low values across all models (maximum 0.666, cf. Table 3).

Influence of training data volume

Increasing training data volume did not consistently improve performance (Table 3). In several cases, models trained on fewer, but manually cut segments outperformed models trained on larger datasets. However, when the well-curated data were incorporated into the full dataset, the model trained on the largest dataset (Model Raw) achieved the highest ROC-AUC values for all but two species.

Comparison of classifier performance

Mean ROC-AUC and PR-AUC values suggested a performance advantage of our AnuraPAM Raw model over FrogNET and Perch v2 (Table 4). Detailed species-specific performance metrics can be found in Supplementary document 1.

Availability and application

All our four custom classifiers are compiled in the AnuraPAM package and are available in both available out-

Table 4. Mean performance metrics of AnuraPAM model (Raw) compared to FrogNET and Perch v2; highest and lowest values per metric are in bold and italic characters, respectively.

Model	AnuraPAM (Raw)	FrogNET	Perch v2
Mean ROC-AUC	0,973	0,936	<i>0,906</i>
Mean PR-AUC	0,880	0,800	<i>0,793</i>

put formats of BirdNET-Analyzer (tflite, raven) in Zenodo (<https://doi.org/10.5281/zenodo.20610756>). This makes them suitable for use with BirdNET-Analyzer and other freely available software, such as PAMalyzer and Raven Intelligence. A user guide explaining the implementation steps can be found in Supplementary document 1.

We recommend AnuraPAM model Raw for general applications as it achieves the highest average ROC-AUC and PR-AUC values. The alternative models should only be used in cases, in which a higher performance of specific focal taxa is desirable (e.g., target-species surveys). The practical applicability of all our custom-trained classifiers would be substantially improved by deployment in a mobile application. As no such application is currently available, their use is limited to devices that support software compatible with the models' file formats. The models can be applied to recordings from various sampling schemes, including long-term passive acoustic monitoring at fixed sites and smartphone recordings from multiple locations. Nevertheless, context-specific manual validation is recommended, since misclassifications may occur in novel acoustic environments.

Discussion

The results demonstrate that automated acoustic identification of Central European anurans using BirdNET is feasible and reliable for most species. The performance levels of our AnuraPAM classifiers are also comparable to those reported in other CNN-based bioacoustic studies (GHANI et al. 2025, GHANI et al. 2023, JACUZZI & OLDEN 2025, KNIGHT et al. 2017, STOWELL et al. 2019).

Our results highlight the capabilities of custom classifiers trained on a relatively small amount of well-curated data. However, hybrid taxa, particularly *Pelophylax kl. esculentus*, remain difficult to classify due to biological and acoustic overlap. The high ROC-AUC values obtained for most species confirm that convolutional neural networks are well suited for automated amphibian call classification, even in acoustically complex freshwater environments. Recent comparative studies across taxa have similarly demonstrated that CNN-based classifiers outperform traditional machine-learning approaches when applied to bio-

acoustic datasets, particularly under variable noise conditions (DIAS et al. 2022). Our results therefore align with a growing consensus that deep-learning approaches represent the current state of the art in passive acoustic monitoring. The suitability of ROC-based metrics for performance evaluation of classifiers with imbalanced datasets is considered imprecise. Previous studies suggest that the high number of true negatives in imbalanced datasets can mask a poor detection rate of minority classes (DAVIS & GOAD-RICH 2006, SAITO & REHMSMEIER 2015, VAN MERRIËN-BOER et al. 2024). Therefore, PR-based metrics are usually recommended to evaluate performance of classifiers with unbalanced datasets. However, RICHARDSON et al. (2024) reported that ROC-based metrics can provide a robust assessment even for imbalanced datasets, as long as changing the imbalance does not affect the score distribution. To accord to both findings we provided the ROC-AUC values, as well as the PR-AUC values in this study. This approach enables comparability to previous and future studies that may only use one of these metrics.

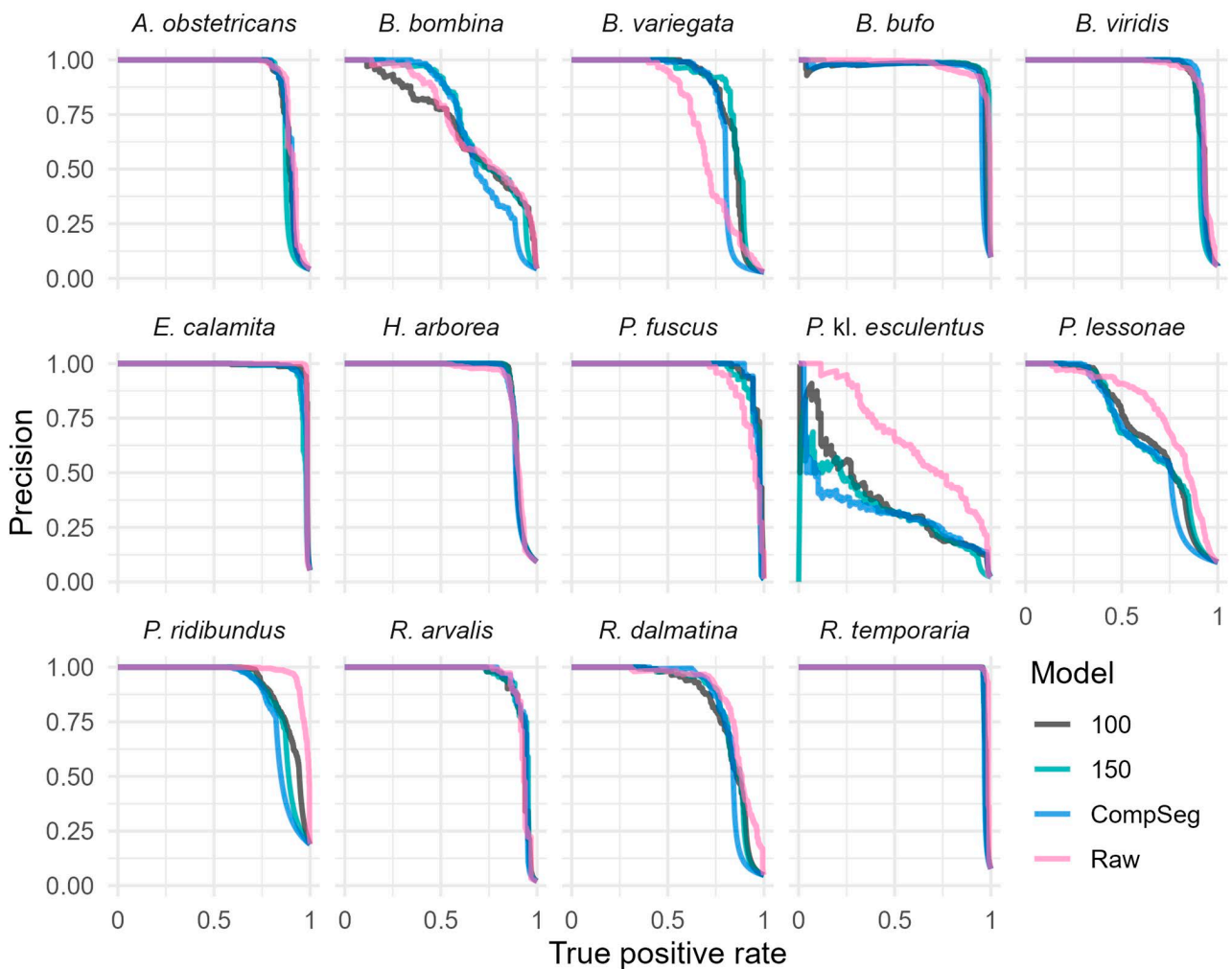


Figure 2. PR curves per species of the four different AnuraPAM models evaluated on an independent test data set.

The consistently low classification performance for *Pelophylax* kl. *esculentus* reflects a broader challenge in automated bioacoustics: the classification of hybrid or cryptic taxa. Recent studies emphasize that hybrids often exhibit highly variable and context-dependent call characteristics, limiting the effectiveness of purely acoustic classifiers (DE ALMEIDA et al. 2024). This issue is particularly relevant in conservation contexts, as hybrid taxa may play distinct ecological roles or be subject to different legal protections. The situation is further compounded by the ongoing invasion of native water frog lineages and their hybridogenic hybrids by alien water frog lineages, largely facilitated by the frog-leg industry and the pet trade, resulting in substantial genetic introgression (DUFRESNES et al. 2024). This has, for example, been shown for western European populations of *P. lessonae* and their hybrid *P. kl. esculentus* invaded by the Italian pool frog *P. l. bergeri* (DUFRESNES & DUBEY 2020). Such introgression alters both the genetic constitution of local populations and their acoustic characteristics, yet its implications remain underexplored in bioacoustics studies. Our findings suggest that reliable automated detection of *P. kl. esculentus* will likely require substantially expanded, expert-curated datasets that use call recordings of genotyped individuals/populations and/or multimodal approaches combining acoustics with genetic/genomic or environmental data. Recent work on multimodal biodiversity monitoring indicates that combining PAM with environmental DNA (eDNA) could be a useful but context-dependent strategy (CHEN et al. 2023).

One of the findings of this study is that increasing training data volume alone does not necessarily improve classification performance. Poorly curated or acoustically redundant data might lead to overfitting, reducing a model's ability to generalize to new environments. As such, emphasis should be placed on the curation and choice of training data. As reported in KATH et al. (2024) it is important to actively choose training segments which provide the models with information that gives them the best chance to reduce false classification during training. In this study we selected only high-quality recordings for all of the models and validated species occurrence before training. Although models 100, 150, and CompSeg were trained on only a subset of the available recordings, their performance is comparable to Model Raw as well as Perch v2 and FrogNET, which were trained on substantially larger datasets. This suggests that, when carefully curated, even a relatively small dataset can be sufficient to train a robust and reliable classification model.

Policy relevance and methodological implications

Standardized and harmonized biodiversity monitoring approaches are increasingly indispensable for fulfilling international and European reporting obligations, including those arising from the EU Habitats Directive and the Convention on Biological Diversity (CBD), e.g., the 2030 Global Biodiversity Framework (MOERSBERGER et al. 2024). Although we have seen a steady increase in monitoring initia-

tives over the past decade, the current rate and scale of global biodiversity loss exceeds our ability to measure, monitor and predict trends using existing biodiversity monitoring approaches (GONZALEZ et al. 2023), making (semi-)automated monitoring approaches even more important. Automated acoustic classifiers facilitate standardized and repeatable surveys across spatial and temporal scales that are difficult to achieve using conventional field methods. In particular, passive acoustic monitoring (PAM) combined with convolutional neural networks has proven well suited for detecting changes in calling phenology, activity patterns, and long-term population trends, which are of growing relevance in the context of climate change (SUGAI et al. 2019). Automated acoustic monitoring of amphibians is directly applicable to biodiversity conservation policy in Germany and the European Union. Article 11 of the Habitats Directive (92/43/EEC) obliges Member States to establish surveillance systems for assessing the conservation status of species and habitat types of community interest, while Article 17 requires six-yearly reporting on distribution, population trends, habitat quality, and future prospects. Anuran species listed in Annexes II and IV, including nine out of the 14 species analyzed and detected with high precision in this study (*A. obstetricans*, *B. bombina*, *B. variegata*, *E. calamita*, *H. arborea*, *P. fuscus*, *P. lessonae* and *R. arvalis*) are of particular relevance in this context. The BirdNET-based classifier presented here offers a complementary tool to support the EU Habitats Directive surveillance obligations. PAM enables standardized data collection over extended temporal periods and repeated sampling at fixed sites, which is essential for detecting population trends and phenological shifts relevant to Article 17 assessments. Automated classification further reduces observer bias and personnel demands, thereby increasing the feasibility of long-term monitoring programmes. Importantly, automated approaches are consistent with recent recommendations to integrate technological innovations into official monitoring schemes in order to improve cost efficiency and data completeness (ROSS et al. 2023, BROWNING et al. 2017). Although such tools cannot replace expert-based field surveys, they can substantially enhance spatial and temporal resolution and function as early-warning systems for population declines. Our results highlight the necessity of species-specific validation, especially for acoustically similar taxa and hybrid systems such as the *Pelophylax* complex. For species of the EU Habitats Directive, conservative classification thresholds and expert validation of critical records are advisable to minimize the risk of erroneous reporting. Automated tools should be embedded within adaptive monitoring frameworks that combine technological efficiency with expert oversight (SUGAI et al. 2019).

Citizen science and participatory monitoring

Citizen science plays an increasingly important role in biodiversity data collection in Europe, contributing substantially to distribution atlases and Red List assessments. However, its integration into EU Habitats Directive report-

ing remains limited, primarily due to concerns regarding data quality and standardization. However, recent initiatives to monitor amphibians show that this widespread scepticism is not always justified and is often based on a lack of knowledge about the potential of citizen science projects (LEONHARDT et al. 2026). Automated acoustic classifiers in particular can help mitigate these concerns by enabling standardized species identification and automated quality control. The widespread public use of BirdNET lowers technical barriers to participation and offers opportunities to incorporate citizen-collected acoustic data into EU Habitats Directive surveillance under Article 11, particularly for common but under-monitored species and the anuran call classifier introduced here provides a potentially valuable addition to this toolkit.

Perspectives and outlook

Future improvements of the trained AnuraPAM classifiers should primarily focus on expanding and curating the training data base and on targeted methodological refinements. Increasing sample sizes for taxa with limited classification performance has been shown to improve model reliability, provided that high annotation quality is maintained (SYMES et al. 2023). Annotation accuracy can be enhanced through multi-observer validation or dependent-observer approaches, which reduce identification errors and increase precision (DRAKE et al. 2016, FORCEY et al. 2006). In particular, the *Pelophylax* complex would likely benefit from a more comprehensive and expert-curated, ideally genotypically pre-validated, data base, as classification performance varied strongly among taxa and was especially low for *P. kl. esculentus*. Recent work demonstrates that deep-learning approaches can successfully separate *Pelophylax* species and hybrids acoustically when trained on suitable data (DE ALMEIDA et al. 2024). Methodologically, replacing the one-vs-rest approach with a one-vs-one framework may allow a more detailed, pairwise assessment of class separability and facilitate targeted model optimization (VANACORE et al. 2024, MEHRA & GUPTA 2013). Although extending the set of evaluation metrics (e.g., F-score) can aid interpretation, no single metric fully captures model performance, and metric choice should therefore remain context-dependent (KNIGHT et al. 2017, VAN MERRIËNBOER et al. 2024). Given the high intra- and interspecific variability of anuran vocalizations (KÖHLER et al. 2017, YUE et al. 2017, TOLEDO et al. 2015), future data sets should aim to represent the full spectrum of call types across geographic regions to ensure robust classification (OSWALD et al. 2022). In addition, assembling representative, non-random test data sets may provide more realistic estimates of model transferability (MUTANU et al. 2022). Model development should thus be regarded as an iterative trial-and-error process in which methodological and data-driven adjustments are continuously evaluated (DESPREZ et al. 2023, BRAUER et al. 2016). In applied contexts such as conservation monitoring, automated classifiers offer substantial potential to reduce time and cost, provided that

users are trained in result interpretation and model limitations are clearly communicated (ROSS et al. 2023, OSWALD et al. 2022). For field applications, models showing consistently high performance across most taxa are suitable, whereas results for the *Pelophylax* complex should be interpreted with caution and supplemented by manual verification. Ensuring high-quality input data and adapting models to local soundscapes, including noise classes, will further enhance practical performance (MADHUSUDHANA et al. 2022). In this context, few-shot transfer learning represents a promising and efficient strategy for locally adapting classifiers with limited additional data (JACUZZI & OLDEN 2025). The AnuraPAM call classifiers represent a freely available, ready-to-use alternative to existing approaches, combining robust performance with practical flexibility and adaptability. By integrating AnuraPAM into standardized platforms such as PAMalyzer, we intend to lower technical barriers and facilitate immediate application in both research and conservation contexts. AnuraPAM can also serve as a foundation for future anuran call classification models that use emerging methods such as ensemble learning (LU et al. 2026). We explicitly encourage broad and intensive adoption of the AnuraPAM classifiers, as their routine use has the potential to substantially expand the volume and diversity of acoustic monitoring data. Such increased deployment will not only enhance conservation practice but also strengthen standard amphibian monitoring programs by generating consistent, large-scale datasets that enable continuous, data-driven refinement of classifier performance.

Note added in proof

A very easy way to use AnuraPAM for amphibian sound classification in the field is the BatGizmo App for Android by John Mears. Install the app (<https://twilighttravels.org/batgizmo-app/>) on your smartphone, download the “AnuraPAM-classifier for BatGizmo” from the Zenodo archive, import the classifier into the app and start to classify amphibians in the field.

Acknowledgements

We would like to thank JÖRN KÖHLER and an anonymous reviewer for their comments and suggestions, which significantly improved the manuscript. We sincerely thank KURT GROSSENBACHER for granting access to his exceptional anuran call collection and for valuable discussions regarding these data. We are also grateful to AXEL KWET and BENNY TRAPP for providing additional call recordings. In particular, we thank A. KWET for drawing our attention to K. GROSSENBACHER’s unique collection. We owe our gratitude to all contributors who have made their personal recordings available through citizen science platforms such as Xeno-canto and the Sonothèque. Without their willingness to share these data, this work would not have been possible. The publication of this article is funded by the Open Access Publication Fund of Hochschule für Technik und Wirtschaft Dresden – University of Applied Sciences. FLORIAN MEERHEIM is supported by the European Union through the European Social Fund Plus (ESF+) within the framework of the ESF PLUS programme “Cooperative State Innovation doctoral candidates” and the State of Saxony.

References

- ANUNCIAÇÃO, P. R., L. S. M. SUGAI, F. MARTELLO, L. M. T. DE CARVALHO & M. C. RIBEIRO (2022): Estimating the diversity of tropical anurans in fragmented landscapes with acoustic monitoring: lessons from a sampling sufficiency perspective. – *Biodiversity and Conservation*, **31**: 3055–3074.
- BANKS-LEITE, C., R. M. EWERS, H. FOLKARD-TAPP & A. FRASER (2020): Countering the effects of habitat loss, fragmentation, and degradation through habitat restoration. – *One Earth*, **3**: 672–676.
- BARNOSKY, A. D., N. MATZKE, S. TOMIYA, G. O. U. WOGAN, B. SWARTZ, T. B. QUENTAL, C. MARSHALL, J. L. MCGUIRE, E. L. LINDSEY, K. C. MAGUIRE, B. MERSEY & E. A. FERRER (2011): Has the Earth's sixth mass extinction already arrived? – *Nature*, **471**: 51–57.
- BRAUER, C. L., T. M. DONOVAN, R. M. MICKY, J. KATZ & B. R. MITCHELL (2016): A comparison of acoustic monitoring methods for common anurans of the northeastern United States. – *Wildlife Society Bulletin*, **40**: 140–149.
- BROWNING, E., R. GIBB, P. GLOVER-KAPFER & K. E. JONES (2017): Passive acoustic monitoring in ecology and conservation. – *WWF Conservation Technology Series*, WWF-UK, Woking.
- CHEN, Y., O. TOURNAYRE, H. TIAN & S. C. LOUGHEED (2023): Assessing the breeding phenology of a threatened frog species using eDNA and automatic acoustic monitoring. – *PeerJ*, **11**: e14679.
- COGĂLNICEANU, D., F. STĂNESCU & C. EGGERT (2024): Aging and growth outside the comfort zone: the case of alien fire-bellied toads in France. – *European Zoological Journal*, **91**: 142–150.
- COLE, J. S., N. L. MICHEL, S. A. EMERSON & R. B. SIEGEL (2022): Automated bird sound classifications of long-duration recordings produce occupancy model outputs similar to manually annotated data. – *Ornithological Applications*, **124**: duacoo3.
- COLONNA, J., T. PEET, C. A. FERREIRA, A. M. JORGE, E. F. GOMES & J. GAMA (2016): Automatic classification of anuran sounds using convolutional neural networks. – pp. 73–78 in: DESAI, E. (ed.): *Proceedings of the Ninth International C* Conference on Computer Science & Software Engineering*. – Association for Computing Machinery, Porto.
- CREEMERS, R. & J. VAN DELFT (2009): *Nederlandse Fauna 9 – De amfibieën en reptielen van Nederland*. – Nationaal Natuurhistorisch Museum Naturalis & European Invertebrate Survey Netherlands, Leiden.
- DAVIS, J. & M. GOADRICH (2006): The relationship between Precision-Recall and ROC curves. – pp. 233–240 in: COHEN, W. W. & A. MOORE (eds): *Proceedings of the 23rd International Conference on Machine Learning*. – Association for Computing Machinery, Pittsburgh.
- DE ALMEIDA, F. G., F. CIAVATTI, A. PULIDO, T. GENDRE, P. GENIEZ, A. EBLE, E. LEBLANC, Y. MANSIER, I. MARTINOSI-ALLIBERT, B. RAYMOND, A. PINEAU, J. DEMAY & P.-A. CROCHET (2024): Acoustics of parental and hybridogenetic water frogs (*Pelophylax*: Amphibia: Ranidae): a tool for monitoring an invasive species. – *Amphibia-Reptilia*, **45**: 333–348.
- DESPREZ, M., V. MIELE & O. GIMENEZ (2023): Nine tips for ecologists using machine learning. – *arXiv*, **2305**:10472.
- DIAS, F. F., M. A. PONTI & R. MINGHIM (2022): A classification and quantification approach to generate features in sound-scape ecology using neural networks. – *Neural Computing and Applications*, **34**: 1923–1937.
- DRAKE, K. L., M. FREY, D. HOGAN & R. HEDLEY (2016): Using digital recordings and sonogram analysis to obtain counts of yellow rails. – *Wildlife Society Bulletin*, **40**: 346–354.
- DUFRESNES, C. & S. DUBEY (2020): Invasion genomics supports an old hybrid swarm of pool frogs in western Europe. – *Biological Invasions*, **22**: 205–210.
- DUFRESNES, C., B. MONOD-BROCA, A. BELLATI, D. CANESTRELLI, J. AMBU, B. WIELSTRA, S. DUBEY, P.-A. CROCHET, M. DENOËL, & D. JABLONSKI (2024): Piecing the barcoding puzzle of Palearctic water frogs (*Pelophylax*) sheds light on amphibian biogeography and global invasions. – *Global Change Biology*, **30**: e17180.
- FICETOLA, G. F., C. COÏC, M. DETAINT, M. BERRONEAU, O. LORVELEC & C. MIAUD (2007): Pattern of distribution of the American bullfrog *Rana catesbeiana* in Europe. – *Biological Invasions*, **9**: 767–772.
- FORCEY, G. M., J. T. ANDERSON, F. K. AMMER & R. C. WHITMORE (2006): Comparison of two double-observer point-count approaches for estimating breeding bird abundance. – *Journal of Wildlife Management*, **70**: 1674–1681.
- FROST, D. R. (2026): *Amphibian Species of the World: An Online Reference*. Version 6.2. – American Museum of Natural History, New York, Electronic Database available online at <https://amphibiansoftheworld.amnh.org/index.php>, accessed 17 January 2026.
- GHANI, B., T. DENTON, S. KAHL & H. KLINCK (2023): Global bird-song embeddings enable superior transfer learning for bio-acoustic classification. – *Scientific Reports*, **13**: 22876.
- GHANI, B., V. J. KALKMAN, B. PLANQUÉ, W.-P. VELLINGA, L. GILL & D. STOWELL (2025): Impact of transfer learning methods and dataset characteristics on generalization in birdsong classification. – *Scientific Reports*, **15**: 16273.
- GIBB, R., E. BROWNING, P. GLOVER-KAPFER & K. E. JONES (2019): Emerging opportunities and challenges for passive acoustics in ecological assessment and monitoring. – *Methods in Ecology and Evolution*, **10**: 169–185.
- GONZALEZ, A., P. VIHervaara, P. BALVANERA, A. E. BATES, E. BAYRAKTAPOV, P. J. BELLINGHAM, A. BRUDER, J. CAMPBELL, M. D. CATCHEN, J. CAVENDER-BARES, J. CHASE, N. COOPS, M. J. COSTELLO, M. DORNELAS, G. DUBOIS, E. J. DUFFY, H. EGGERMONT, N. FERNANDEZ, S. FERRIER, G. N. GELLER, M. J. GILL, D. GRAVEL, C. A. GUERRA, R. P. GURALNICK, M. HARFOOT, T. HIRSCH, S. M. HOBAN, A. C. HUGHES, M. HUNTER, F. ISBELL, W. JETZ, N. JUERGENSEN, W. D. KISSLING, C. B. KRUG, Y. LE BRAS, B. LEUNG, M. C. LONDOÑO-MURCIA, J.-M. LORD, M. LOREAU, A. LUERS, K. MA, A. J. MACDONALD, M. MCGEOCH, K. L. MILLETTE, Z. MOLNAR, A. S. MORI, F. E. MULLER-KARGER, H. MURAOKA, L. M. NAVARRO, T. NEWBOLD, A. NIAMIR, D. OBURO, M. O'CONNOR, M. PAGANINI, H. M. PEREIRA, T. POISOT, L. J. POLLOCK, A. PURVIS, A. RADULOVICI, D. ROCCHINI, M. E. SCHAEPMAN, G. SCHAEPMAN-STRUB, D. S. SCHMELLER, U. SCHMIEDEL, F. D. SCHNEIDER, M. M. SHAKYA, A. K. SKIDMORE, A. L. SKOWNO, Y. TAKEUCHI, M.-N. TUANMU, E. TURAK, W. TURNER, M. C. URBAN, N. URBINA-CARDONA, R. VALBUENA, B. VAN HAVRE & E. WRIGHT (2023): A global biodiversity observing system to unite monitoring and guide action. – *Nature Ecology & Evolution*, **7**: 1947–1952.
- HASSAN, N., D. A. RAMLI & H. JAAFAR (2017): Deep neural network approach to frog species recognition. – pp. 173–178 in: *Institute of Electrical and Electronics Engineers* (ed.): 2017

- IEEE 13th International Colloquium on Signal Processing & its Applications (CSPA). – IEEE, Penang.
- HOFFMANN, A. & H.-U. REYER (2013): Genomic effects on advertisement call structure in diploid and triploid hybrid waterfrogs (*Anura, Pelophylax esculentus*). – *BMC Ecology*, **13**: 47.
- HUANG, C.-J., Y.-J. YANG, D.-X. YANG & Y.-J. CHEN (2009): Frog classification using machine learning techniques. – *Expert Systems with Applications*, **36**: 3737–3743.
- IPBES (2019): Global assessment report on biodiversity and ecosystem services of the Intergovernmental Science-Policy Platform on Biodiversity and Ecosystem Services. – IPBES Secretariat, Bonn.
- IUCN (2026): The IUCN Red List of Threatened Species. Version 2026-1. – IUCN Biodiversity Assessment & Knowledge Team: Red List Unit, Cambridge, available online at <https://www.iucnredlist.org>, accessed 9 June 2026.
- JANSSEN, J. A. M., J. R. RODWELL, M. GARCÍA CRIADO, S. GUBBAY, T. HAYNES, A. NIETO, N. SANDERS, F. LANDUCCI, J. LOIDI, A. SSYMANK, T. TAHVANAINEN, M. VALDERRABANO, A. ACOSTA, M. ARONSSON, G. ARTS, F. ATTORRE, E. BERGMEIER, R.-J. BIJLSMA, F. BIORET, C. BIŢĂ-NICOLAE, I. BIURRUN, M. CALIX, J. CAPELO, A. ČARNI, M. CHYTRÝ, J. DENGLE, P. DIMOPOULOS, F. ESSL, H. GARDFJELL, D. GIGANTE, G. GIUSO DEL GALDO, M. HÁJEK, F. JANSEN, J. JANSEN, J. KAPFER, A. MICKOLAJCZAK, J. A. MOLINA, Z. MOLNÁR, D. PATERNOSTER, A. PIERNIK, B. POULIN, B. RENAUX, J. H. J. SCHAMINÉE, K. ŠUMBEROVÁ, H. TOIVONEN, T. TONTERI, I. TSIRIPIDIS, R. TZONEV & M. VALACHOVIČ (2016): European red list of habitats. Part 2, Terrestrial and freshwater habitats. – Publications Office of the European Union, Luxembourg.
- JACUZZI, G. & J. D. OLDEN (2025): Few-shot transfer learning enables robust acoustic monitoring of wildlife communities at the landscape scale. – *Ecological Informatics*, **90**: 103294.
- KAHL, S., C. M. WOOD, M. EIBL & H. KLINCK (2021): BirdNET: A deep learning solution for avian diversity monitoring. – *Ecological Informatics*, **61**: 101236.
- KATH, H., P. P. SERAFINI, I. B. CAMPOS, T. S. GOUVÊA & D. SONNTAG (2024): Leveraging transfer learning and active learning for data annotation in passive acoustic monitoring of wildlife. – *Ecological Informatics*, **82**: 102710.
- KIRKLAND, M. (2025): FrogNET v1.0.2. – available online at <https://chirpity.net/downloads/FrogNET%20v1.0.2.zip>, accessed 9 June 2026.
- KNIGHT, E. C., K. C. HANNAH, G. J. FOLEY, C. D. SCOTT, R. M. BRIGHAM & E. BAYNE (2017): Recommendations for acoustic recognizer performance assessment with application to five common automated signal recognition programs. – *Avian Conservation and Ecology*, **12**: 14.
- KÖHLER, J., M. JANSEN, A. RODRÍGUEZ, P. J. R. KOK, L. F. TOLEDO, M. EMMRICH, F. GLAW, C. F. B. HADDAD, M.-O. RÖDEL & M. VENCES (2017): The use of bioacoustics in anuran taxonomy: theory, terminology, methods and recommendations for best practice. – *Zootaxa*, **4251**: 1–124.
- KRIZMANIĆ, I. I. & A. T. IVANOVIĆ (2010): Population systems of the *Pelophylax esculentus* complex in the southern part of its range. – *Folia Zoologica*, **59**: 215–222.
- LEONHARDT, F., F. MEERHEIM, G. L. NACCARATO, R. WALZ & R. ERNST (2026): Vergangenheit trifft Zukunft – Historische Daten und Citizen Science erhöhen die Nachweisdichte des Feuersalamanders in der Sächsischen Schweiz. – *Natur und Landschaft*, **101**: 62–71.
- LU, J., Z. WANG, M. SI, L. GONG, J. FENG & T. JIANG (2026): Deep ensemble learning for automated acoustic identification of bats: A case study on the SCAR dataset. – *Expert Systems with Applications*, **317**: 132016.
- MADHUSUDHANA, S., G. PAVAN, L. A. MILLER, W. L. GANNON, A. HAWKINS, C. ERBE, J. A. HAMEL & J. A. THOMAS (2022): Choosing equipment for animal bioacoustic research. – pp. 37–85 in: ERBE, C. & J. A. THOMAS (eds): *Exploring Animal Behavior Through Sound*. – Springer International Publishing, Cham.
- MARSLAND, S., N. PRIYADARSHANI, J. JUODAKIS & I. CASTRO (2019): AviaNZ: A future-proofed program for annotation and recognition of animal sounds in long-time field recordings. – *Methods in Ecology and Evolution*, **10**: 1189–1195.
- MEHRA, N. & S. GUPTA (2013): Survey on multiclass classification methods. – *International Journal of Computer Science and Information Technologies*, **4**: 572–576.
- MOERSBERGER, H., J. VALDEZ, J. G. MARTIN, J. JUNKER, I. GEORGIEVA, S. BAUER & A. BONN (2024): Biodiversity monitoring in Europe: User and policy needs. – *Conservation Letters*, **17**: e13038.
- Muséum national d'Histoire naturelle (2026): Sonothèque. – Available online at <https://sonotheque.mnhn.fr/>, accessed 9 June 2026.
- MUTANU, L., J. GOHIL, K. GUPTA, P. WAGIO & G. KOTONYA (2022): A review of automated bioacoustics and general acoustics classification research. – *Sensors*, **22**: 8361.
- OSWALD, J. N., A. M. VAN CISE, A. DASSOW, T. ELLIOTT, M. T. JOHNSON, A. RAVIGNANI & J. PODOS (2022): A collection of best practices for the collection and analysis of bioacoustic data. – *Applied Sciences*, **12**: 12046.
- RICHARDSON, E., R. TREVIZANI, J. A. GREENBAUM, H. CARTER, M. NIELSEN & B. PETERS (2024): The receiver operating characteristic curve accurately assesses imbalanced datasets. – *Patterns*, **5**: 100994.
- ROSS, S. R.-J., D. P. O'CONNELL, J. L. DEICHMANN, C. DESJONQUÈRES, A. GASC, J. N. PHILLIPS, S. S. SETHI, C. M. WOOD & Z. BURIVALOVA (2023): Passive acoustic monitoring provides a fresh perspective on fundamental ecological questions. – *Functional Ecology*, **37**: 959–975.
- Rote-Liste-Gremium Amphibien und Reptilien (2020): Rote Liste und Gesamtartenliste der Amphibien (Amphibia) Deutschlands. – Naturschutz und Biologische Vielfalt, **170**. Bundesamt für Naturschutz, Bonn-Bad Godesberg.
- SAITO, T. & M. REHMSMEIER (2015): The precision-recall plot is more informative than the ROC plot when evaluating binary classifiers on imbalanced datasets. – *PloS One*, **10**: e0118432.
- SAYER, C. A., E. FERNANDO, R. R. JIMENEZ, N. B. W. MACFARLANE, G. RAPACCIUOLO, M. BÖHM, T. M. BROOKS, T. CONTRERAS-MACBEATH, N. A. COX, I. HARRISON, M. HOFFMANN, R. JENKINS, K. G. SMITH, J.-C. VIÉ, J. C. ABBOTT, D. J. ALLEN, G. R. ALLEN, V. BARRIOS, J.-P. BOUDOT, S. F. CARRIZO, P. CHARVET, V. CLAUSNITZER, L. CONGIU, K. A. CRANDALL, N. CUMBERLIDGE, A. CUTTELOD, J. DALTON, A. G. DANIELS, S. DE GRAVE, G. DE KNIJF, K.-D. B. DIJKSTRA, R. A. DOW, J. FREYHOF, N. GARCÍA, J. GESSNER, A. GETAHUN, C. GIBSON, M. J. GOLLOCK, M. I. GRANT, A. E. R. GROOM, M. P. HAMMER, G. A. HAMMERSON, C. HILTON-TAYLOR, L. HODGKINSON, R.

Supplementary data

The following data are available online:

Supplementary document 1. Calculated ROC-AUC and PR-AUC values per species for the AnuraPAM models trained (model 100, model 150, model CompSeg (CS), model Raw) and 2 available models (FrogNET, Perch v2); Guidelines for the Practical Application of AnuraPAM models.

- A. HOLLAND, R. W. JABADO, D. JUFFE BIGNOLI, V. J. KALKMAN, B. K. KARIMOV, J. KIPPING, M. KOTTELAT, P. A. LALÈYÈ, H. K. LARSON, M. LINTERMANS, F. LOZANO, A. LUDWIG, T. J. LYONS, L. MÁIZ-TOMÉ, S. MOLUR, H. H. NG, C. NUMA, A. F. PALMER-NEWTON, C. PIKE, H. E. PIPPARD, C. N. M. POLAZ, C. M. POLLOCK, R. RAGHAVAN, P. S. RAND, T. RAVELOMANANA, R. E. REIS, C. L. RIGBY, J. A. SCOTT, P. H. SKELTON, M. R. SLOAT, J. SNOEKS, M. L. J. STIASSNY, H. H. TAN, Y. TANIGUCHI, E. B. THORSTAD, M. F. TOGNETTI, A. G. TORRES, Y. TORRES, D. TWEDDLE, K. WATANABE, J. R. S. WESTRIE, E. G. E. WRIGHT, E. ZHANG & W. R. T. DARWALL. (2025): One-quarter of freshwater fauna threatened with extinction. – *Nature*, **638**: 138–145.
- SCHNEIDER, H. (2005): Bioakustik der Froschlurche: Einheimische und Verwandte Arten. – *Supplemente der Zeitschrift für Feldherpetologie*, **6**. – Laurenti-Verlag, Bielefeld.
- SILLERO, N., J. CAMPOS, A. BONARDI, C. CORTI, R. CREEMERS, P.-A. CROCHET, J. C. ISAILOVIĆ, M. DENOËL, G. F. FICETOLA, J. GONÇALVES, S. KUZMIN, P. LYMBERAKIS, P. DE POUS, A. RODRÍGUEZ, R. SINDACO, J. SPEYBROECK, B. TOXOPEUS, D. R. VIEITES & M. VENCES (2014): Updated distribution and biogeography of amphibians and reptiles of Europe. – *Amphibia-Reptilia*, **35**: 1–31.
- STOWELL, D., M. D. WOOD, H. PAMUŁA, Y. STYLIANOU & H. GLOTTIN (2019): Automatic acoustic detection of birds through deep learning: The first Bird Audio Detection challenge. – *Methods in Ecology and Evolution*, **10**: 368–380.
- SUGAI, L. S. M., T. S. F. SILVA, J. W. RIBEIRO & D. LLUSIA (2019): Terrestrial passive acoustic monitoring: Review and perspectives. – *BioScience*, **69**: 15–25.
- SYMES, L., L. S. SUGAI, B. GOTTESMAN, M. PITZRICK & C. WOOD (2023): Acoustic analysis with BirdNET and (almost) no coding: practical instructions (Version 1.0). – Zenodo, <https://doi.org/10.5281/zenodo.8415090>
- TEMPLE, H. J. & N. A. COX (2009): European red list of amphibians. – Office for Official Publications of the European Communities, Luxembourg.
- TETZLAFF, I. (2007): Froschlurche: die Stimmen aller heimischen Arten. – Edition Ample, Germering.
- TOLEDO, L. F., I. A. MARTINS, D. P. BRUSCHI, M. A. PASSOS, C. ALEXANDRE & C. F. B. HADDAD (2015): The anuran calling repertoire in the light of social context. – *Acta Ethologica*, **18**: 87–99.
- VAN MERRIËNBOER, B., J. HAMER, V. DUMOULIN, E. TRIANTAFILOU & T. DENTON (2024): Birds, bats and beyond: evaluating generalization in bioacoustics models. – *Frontiers in Bird Science*, **3**: 1369756.
- VAN MERRIËNBOER, B., V. DUMOULIN, J. HAMER, L. HARRELL, A. BURNS & T. DENTON (2026): Perch 2.0: The bittern lesson for bioacoustics. – arXiv:2508.04665, <https://doi.org/10.48550/arXiv.2508.04665>
- VANACORE, A., M. S. PELLEGRINO & A. CIARDIELLO (2024): Fair evaluation of classifier predictive performance based on binary confusion matrix. – *Computational Statistics*, **39**: 363–383.
- WWF (2024): Living planet report 2024: A system in peril. – WWF, Gland.
- Xeno-canto Foundation (2026): Xeno-Canto: Sharing wildlife sounds from around the world. – available online at <https://xeno-canto.org/>, accessed 9 June 2026.
- YUE, X., Y. FAN, F. XUE, S. E. BRAUTH, Y. TANG & G. FANG (2017): The first call note plays a crucial role in frog vocal communication. – *Scientific Reports*, **7**: 10128.