A Comparative Study of Computational Intelligence Methods for Audio Analysis in Animal Identification within Tropical Ecosystems

Maria J. Guerrero
SISTEMIC, Engineering Faculty
Universidad de Antioquia UdeA
Calle 67 No.53-108
Medellín, Colombia
email: mariaj.guerrero@udea.edu.co

Santiago Taborda
SISTEMIC, Engineering Faculty
Universidad de Antioquia UdeA
Calle 67 No.53-108
Medellín, Colombia
email: santiago.taborda2@udea.edu.co

Juan M. Daza
GHA, Biology Institute
Universidad de Antioquia UdeA
Calle 67 No.53-108
Medellín, Colombia
email: juanm.daza@udea.edu.co

Claudia Isaza
SISTEMIC, Engineering Faculty
Universidad de Antioquia UdeA
Calle 67 No.53-108
Medellín, Colombia
email: victoria.isaza@udea.edu.co

Abstract—Passive Acoustic Monitoring (PAM) using computational intelligence techniques offers new avenues for biodiversity conservation, particularly in identifying and monitoring species within tropical ecosystems. While various methods exist for animal sound identification, a comprehensive understanding of their advantages and disadvantages is often lacking. This work evaluates five methods for automatically identifying species vocalizations across different taxonomic groups using an acoustic dataset from a Colombian agricultural ecosystem. We conducted a comparative analysis of supervised techniques, including Convolutional Neural Networks (CNN), Random Forest (RF), and Support Vector Machine (SVM), as well as unsupervised methods such as spectral clustering, DBSCAN, and the Learning Algorithm for Multivariate Data Analysis (LAMDA) 3pi, evaluating their species detection performance through the F1-Score metric. Our research underscores the critical role of methodological selection in achieving accurate species identification. Furthermore, this study advances the understanding of clustering interpretation, illustrating its potential beyond bioacoustic studies. It presents how unsupervised learning techniques can be valuable in scenarios characterized by limited labeled data, common in tropical ecosystems, and high uncertainty regarding the number of clusters obtained. This approach facilitates the exploration of prototype patterns, aiding species association and potentially extending to other areas requiring insight into unidentified clusters. This study offers valuable insights into selecting suitable tools for bioacoustic studies, emphasizing the need for comprehensive input preparation for model training. The findings underscore the potential of PAM and computational strategies in furthering biodiversity research and conservation efforts, effectively addressing the challenges of species identification and clustering interpretation.

Keywords-Machine learning, Deep learning, Clustering, Bioacoustics, Soundscape, Species identification.

I. INTRODUCTION

Monitoring ecosystems and their species is crucial for understanding and conserving biodiversity [1]. Traditionally, species identification has been performed through direct observation, which faces significant challenges, such as detecting individuals in densely vegetated areas common in tropical ecosystems. Therefore, it is essential to have alternatives that support this task. Passive Acoustic Monitoring (PAM) emerges as an alternative tool that enables the identification of patterns in ecosystems using sound from different sources, such as biological organisms, geophysical phenomena, and human activities [2]. This alternative involves deploying acoustic sensors to record sounds across different areas, thereby serving as a helpful tool to detect the presence of animal species at a specific location and time, answering biological questions that help identify the conservation state of ecosystems due to external impacts [3].

The evolution of sensor technology has significantly enhanced the collection of acoustic data, allowing sampling for months at a time, recording every minute, resulting in vast volumes of acoustic recordings that must be processed and analyzed by biology and ecology experts [4], [5]. In the case of acoustic animal identification, this process, if conducted manually, demands extensive time to identify specific species' vocalizations or calls within each recording. Automatic species call detection algorithms offer a solution by employing various computational intelligence techniques to detect and classify animal vocalizations or calls from diverse taxonomic groups [6]–[11].

Generally, available methodologies encompass a series of stages, such as signal processing, in the case of machine learning methods, segmentation, and feature extraction stages, and finally, the implementation of algorithms to classify species calls. Many of these proposals are implemented in software solutions, such as Avisoft [12], Arbimon [13], and Kaleidoscope Pro [14], offering a practical application of

these methodologies. On the supervised side, commonly used computational intelligence techniques include Convolutional Neural Networks (CNN) [8], [15], [16], Random Forest [17], [18], Support Vector Machine [18], [19], and Hidden Markov Models [12]. Meanwhile, in the unsupervised domain, techniques such as spectral clustering [14], LAMDA - Learning Algorithm for Multivariate Data Analysis - [10], [16], [20], and DBSCAN [21] are available.

Each technique within species identification brings its own set of characteristics and advantages, tailored to address distinct research questions ranging from species-specific, individual identification to the detection of multiple animal species. While supervised methodologies dominate the field, requiring extensive, labeled datasets for model training [22], unsupervised identification methods are necessary. These approaches, free from the constraints of prior species knowledge or labeled data, offer the potential to uncover data patterns indicative of specific species in a habitat, especially in countries with high biodiversity like Colombia, where unknown species to science still exist, making the process of obtaining large, labeled datasets difficult. However, the challenge with unsupervised techniques, such as clustering, primarily revolves around interpreting the results. Accurately associating each cluster with a particular species requires time and expert knowledge.

This work aims to analyze and compare different machine learning and deep learning methodologies, encompassing supervised and unsupervised approaches, proposed for automatically detecting calls and vocalizations of multiple animal species in soundscapes. Additionally, by leveraging the unsupervised approach proposed in [10], we performed an analysis to interpret clusters and automatically assign them to specific species. The performance of each method was assessed using a collection of audio recordings from the Colombian agricultural ecosystem. Our objective was to facilitate the selection of the most appropriate methodology tailored to the specific research problem, thereby advancing the field of bioacoustic monitoring by identifying the most effective tools for species identification in soundscapes.

The structure of this article is organized as follows: Section II outlines the data utilized, a description of the characteristics of the analyzed algorithms, and a description of clustering interpretation. Section III presents the results obtained from our analysis. Finally, conclusions and future work are drawn in Section IV.

II. MATERIALS AND METHODS

A. Study site and acoustic dataset

The data used in this work were provided by Antioquia's Herpetological Group (GHA), collected via passive acoustic monitoring conducted in a rural area of the municipality of Puerto Wilches, Santander, Colombia ($7°21'52.5"N, 73°51'33.0"W$). The area of this study site is primarily dominated by oil palm plantations, accounting for 75% of the land, of varying ages. Additionally, it encompasses a diverse mixture of secondary vegetation (7.6%), forest patches (6.13%), grasslands (5.5%), and aquatic vegetation

zones (3.2%). The region is dotted with several buildings and crisscrossed by a secondary road network that serves the palm oil and livestock industries.

The analyzed dataset is composed of acoustic data from the audible and ultrasonic spectrum. Audible data consists of a subset of 207 recordings from the primary collection (19,598 recordings), which were rigorously labeled by experts to make the comparison. Experts found 11 species, including six bird species, four anuran species, and one primate. This dataset was collected using a Song Meter Mini device (Wildlife Acoustics, Inc.), configured to record one minute every 10 minutes with a sampling rate of 48 kHz. The ultrasonic dataset consists of 197 recordings collected in the same location as the audible dataset using a Song Meter Mini bat device (Wildlife Acoustics, Inc.), recording 15 s every 15 min with a sampling rate of 384 kHz. In these recordings, 7 species of bats were found. Labels provided by experts were used for the training process of supervised methods and the validation of unsupervised methods.

B. Computational intelligence methods for species identification

The selection of computational intelligence methods for multi-species identification started with a detailed literature review to ascertain the most widely used and popular methodologies in the field. This preliminary phase aims to identify techniques that have been effectively implemented in both software applications and open-source code environments. Further, our analysis examined both supervised and unsupervised learning approaches, ensuring a comprehensive understanding of their application in species identification tasks. The selected approaches are described as follows:

1) Arbimon [13]: Arbimon is a free web-based platform that enhances passive acoustic monitoring with cloud storage and data analysis capabilities. It utilizes artificial intelligence algorithms, such as Random Forest for the supervised classification of animal vocalizations and includes a BETA version of a clustering tool that employs the DBSCAN algorithm for enhanced vocalization detection and categorization. The analysis in this software starts with uploading audio databases to the web platform and organizing them into playlists, which are crucial for the model's training phase. The platform facilitates audio labeling through a pattern matching tool by selecting the acoustic pattern of interest, streamlining the training of the Random Forest model with identified species presences and absences. Additionally, Arbimon is developing an unsupervised tool that includes Acoustic Event Detection (AED) and cluster analysis, using the same database for Random Forest model training. This process requires setting species-specific parameters, such as species frequency information and thresholds, to detect and classify vocalizations accurately. The importance of precise parameter configuration is crucial to achieving accurate detection and analysis outcomes. Each parameter was fine-tuned according to specific species requirements.

2) Raven Pro/Koogu [23]: This methodology integrates the widely recognized Raven Pro software, created by the Cornell Lab of Ornithology [24], known for its spectrogram visualization and manual species calls labeling tools, with the Koogu Python library, designed to train a convolutional neural network model using Raven's selection table. The process involves manually labeling species calls on the Raven platform, organizing species labels and audio recordings, as well as configuring parameters in the Convolutional Neural Network (CNN) model. In this case, a DenseNet architecture was trained, where parameterization included species-specific bandwidth information, the number of training epochs, and batch size. Upon completion of the model's training, the results can be incorporated back into Raven Pro to observe the identification of species calls. In our study, each species was trained individually, allowing for the fine-tuning of parameters to the specific requirements of each species, ensuring precise and effective species identification.

3) Kaleidoscope Pro [14]: This software requires a paid license and is designed to detect animal vocalizations in audible and ultrasound spectrum through a Hidden Markov Model (HMM) and spectral clustering. The use of this software begins with the preparation of an initial database containing audio recordings of the targeted species. This software demands the configuration of several parameters, such as frequency range, the maximum and minimum detection durations, and the maximum time interval between vocalizations, all of which were manually adjusted for each species under study. The clustering analysis option also requires setting the maximum distance to the cluster center, the Fast Fourier Transform (FFT) window size, the number of maximum states, and the maximum cluster number. In this case, after testing with different values, most parameters were maintained at their default settings; the FFT window size was adjusted according to the frequency range of interest (5.33 ms [128 @0–12 kHz, 256 @13–24 kHz, 512 @25–48 kHz, and 1024 @49–96 kHz]).

4) Unsupervised Acoustic Animal Identification [10]: This unsupervised methodology proposed by Guerrero et al. (2023) employs the LAMDA 3pi algorithm to analyze acoustic data without needing pre-defined labels in the database. The methodology simultaneously facilitates analysis across the audible spectrum and ultrasound for multiple species. It includes a segmentation stage that isolates potential acoustic events or species calls, followed by a feature extraction stage that captures relevant acoustic information from each segment. The process culminates in a fuzzy clustering stage, which groups segments based on acoustic similarity, enabling the simultaneous identification of multiple species associated with various taxonomic groups. Although this method effectively identifies vocalizations in all frequency ranges simultaneously and does not require user-defined parameters, it has difficulties with cluster interpretation. The generation of numerous clusters, while beneficial for exhaustive audio analysis, poses a significant challenge for researchers due to the number of clusters to analyze and the inherent intra-cluster variation that results from data uncertainty. This aspect can make the manual examination of each cluster labor-intensive and complex, underscoring a critical consideration in the balance between comprehensive analysis and practical feasibility.

C. Clustering interpretation

Addressing a prevalent challenge in species identification via clustering, we focus on the manual linking between clusters and species-specific vocal patterns. While this method is not constrained by the requirement for labeled data, the process of interpreting clustering results can be time-consuming for experts. We used the fuzzy clustering approach presented in [10], which decomposes the soundscape into acoustic entities named sonotypes (clusters). This method not only validates the possibility of associating sonotypes with species calls but also reveals that the aggregation of these clusters provides insights into acoustic biodiversity patterns over time. Remarkably, this biodiversity pattern aligns with those identified by alternative methods, with the distinctive advantage of enabling the breakdown of these patterns by individual species, a feature unattainable with other techniques.

To streamline the process, our study also ventured into automating the linkage of sonotypes to species by utilizing readily available public datasets or targeted species-specific recordings, aiming to minimize manual efforts in interpreting clustering outcomes.

Pursuing the methodology described in [10], we generated clusters or sonotypes and segmented calls of species detected in the study site utilizing directional recordings from preceding studies and datasets from sources such as xeno-canto [25]. We extracted pivotal time-frequency features from these segments, including minimum and maximum vocalization frequencies, bandwidth, peak frequency, and call duration. These features are also automatically extracted when generating sonotypes. The fuzzy clustering approach enables the determination of each segment's membership degree to its assigned cluster, thereby identifying the cluster's representative element, the segment with the highest degree of belonging to its cluster.

The association of sonotypes to species was then achieved by calculating the Euclidean distance between the median value of the bandwidth, maximum, and minimum frequency bands of the representative elements and the 10 segments with the highest membership values of each sonotype, alongside the same features of manually species segments. Scatter plots illustrating the minimum and maximum frequencies of both species segments and sonotypes aided in visually confirming similarities.

D. Evaluation metrics

In the performance evaluation section, the effectiveness of the algorithms for detecting species calls was assessed using confusion matrices. These matrices were constructed by comparing the original labels from the database against the predictions made by each model, providing values for True Positives (TP), True Negatives (TN), False Positives (FP), and False Negatives (FN). True Positives represent correctly

identified presences, True Negatives denote correctly identified absences, False Positives indicate presences incorrectly classified as absences and False Negatives refer to absences incorrectly classified as presences.

To evaluate model accuracy, especially in cases with unbalanced datasets across different classes, the F1-Score metric is extensively used. This metric, derived from the confusion matrix for each analyzed methodology, offers a balanced measure of a model's precision and recall. The F1-Score is calculated using the following formula:

$$F1 - score = 2 \times \frac{\text{precision} \times \text{recall}}{\text{precision} + \text{recall}} \qquad (1)$$

where precision is defined as the ratio of correctly predicted positive cases to the total predicted positives, indicating the accuracy of positive predictions and recall measures the ratio of actual positives accurately identified by the model, reflecting the model's ability to capture all relevant cases.

III. RESULTS

A. Analysis of computational intelligence methods for species identification

We applied the different methods to our dataset for species identification across different frequency bands, including those within the ultrasound range. For each species, models were constructed using Arbimon in both supervised and unsupervised manners, as well as in the case of Raven Pro/Koogu

and Kaleidoscope Pro, highlighting the general limitations of simultaneously identifying multiple species and species-specific parameterization. Figure 1 showcases the F1-score results achieved with each method, detailing performance across broad taxonomic groups and specific species.

The literature presents several detection proposals for ultrasound species analysis. Within this domain, Kaleidoscope Pro and the approach introduced by Guerrero et al. [10], are distinguishable in their methodologies for effectively identifying both ultrasound and audible species.

Utilizing the Random Forest algorithm, Arbimon exhibited high performance in accurately detecting mammal species and certain bird species. This level of accuracy, however, was not mirrored in the detection of anuran species, with the models notably underperforming in identifying *Leptodactylus fuscus* and *Leptodactylus fragilis*.

The advent of Arbimon's BETA tool, which adopts a cluster analysis via the DBSCAN algorithm, signaled a significant advancement in the identification capabilities for mammals and some avian species, marking an improvement over its supervised analog. Yet, it too faltered with anurans, unable to discern the calls of *Leptodactylus fuscus* and *Leptodactylus fragilis*. Achieving results with this methodology proved challenging due to clustering interpretation. There is no tool to automatically associate clusters with species labels. Additionally, the fine-tuning of detection parameters, such as duration, bandwidth, and threshold areas, became pivotal, as
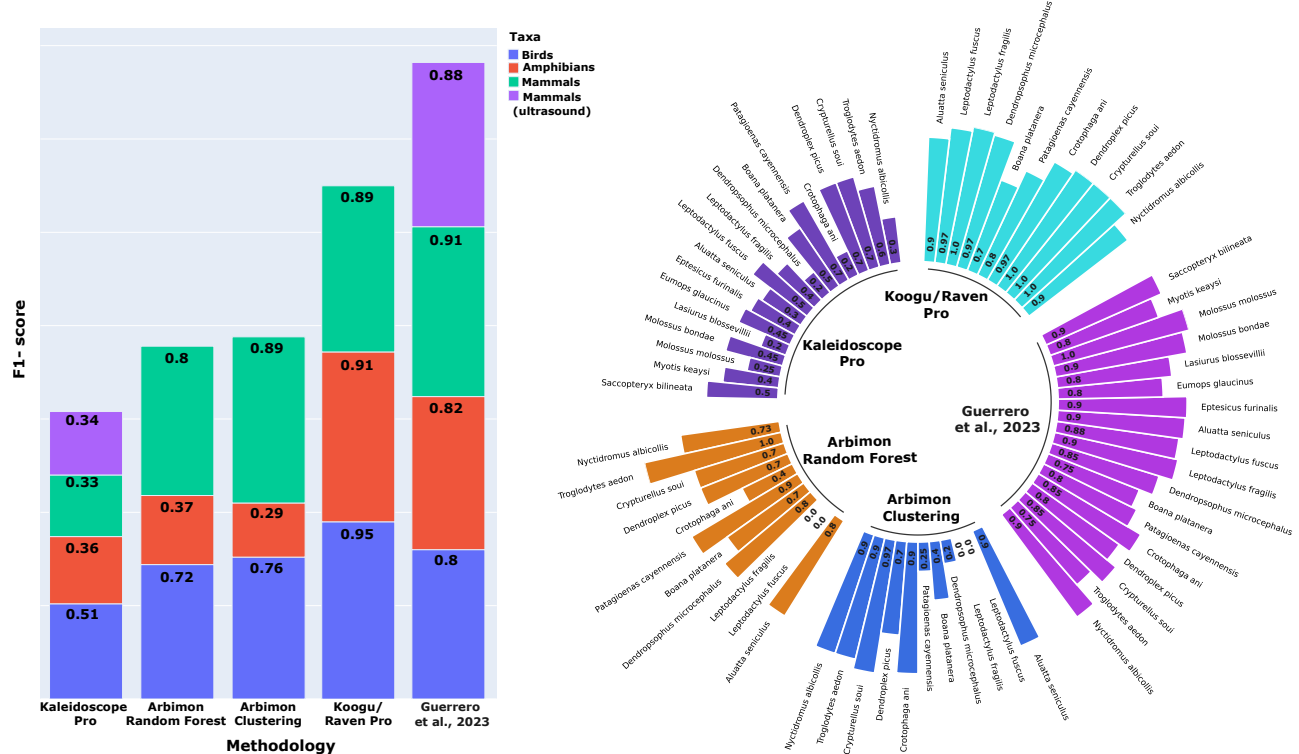

Figure 1. Comparative F1-score results for species identification methodologies. The bar graph on the left delineates the F1-score performance of different methodologies applied to various taxonomic groups. The radial chart on the right provides a detailed breakdown of the F1-scores for each species within these groups.

they significantly influenced the detector's sensitivity to the diverse vocalization characteristics inherent to each species. An improper adjustment of these parameters often resulted in either an overabundance of irrelevant acoustic events or a failure to identify critical species-specific vocal segments, which subsequently compromised the clustering process.

The methodology developed by the Cornell Lab of Ornithology seamlessly combines Raven Pro's capabilities for manual call segmentation and time-frequency feature extraction with the Python library Koogu for generating Convolutional Neural Network (CNN) models. This approach has displayed remarkable efficacy, showcasing high F1-scores across all examined taxonomic groups and most of the species cataloged, thus reinforcing CNNs' versatility and powerful pattern recognition capabilities. CNNs' ability to automatically learn and distinguish subtle differences in spectrogram patterns enables precise species identification from acoustic data.

Koogu is specifically designed to recognize labels from a Raven Pro selection table, making the initiation of the CNN model training straightforward. However, this process faced its own set of challenges, particularly when bandwidth configurations were narrowly aligned with the spectral extremities of segments identified in Raven Pro. To circumvent this constraint, a strategy involving the use of expanded bandwidths was adopted, allowing for a more generous and inclusive detection range.

The performance of Kaleidoscope Pro, in contrast, was found to be less effective than other methods. This tool tended to produce segments of prolonged duration that captured a variety of calls and classified them based on the most prominent pattern in the spectrogram. Consequently, this often led to clusters representing insect stridulations due to their high-intensity frequencies. Moreover, the presence of clusters with

vocalizations from multiple species complicated the process of selecting a representative cluster for each species. While Kaleidoscope Pro provides a manual interface for cluster-to-species association, the task can be complex and prone to inaccuracies when sifting through a vast number of clusters, particularly when some contain a significant amount of noise.

Finally, the approach proposed by Guerrero et al. [10], capable of concurrently detecting species across different frequency bands, showcased F1-scores exceeding 0.75 for all taxonomic groups and species, ultrasound range included. A notable advantage of their fuzzy clustering model is calculating the membership degree of the segments to their assigned clusters. This capability enables further data processing by setting a segment membership threshold. This threshold helps to minimize the number of false positives in the final detections, thereby enhancing the accuracy of species detection. This method also allows for a graphical cluster-to-species association using the acoustic pattern and the sound of the cluster segments. However, similar to Kaleidoscope Pro, this task can be time-consuming if there are a large number of clusters.

B. Clustering interpretation

Implementing the methodology described in Section II-C, our endeavor in clustering interpretation sought to streamline the automatization of associations between species-specific vocalizations and corresponding sonotypes. Within the scope of the audible spectrum study conducted, the LAMDA 3pi algorithm generated 130 clusters, 6 of them successfully associated with 5 out of the 11 species accentuated in Section III-A. This subset of species encompasses three avian species: *Cryptorellus soi*, *Dendroplex picus*, and *Nyctidromus albicollis*, along with two anurans: *Dendropsophus microcephalus* and *Leptodactylus fuscus*. Vocalizations for these species were
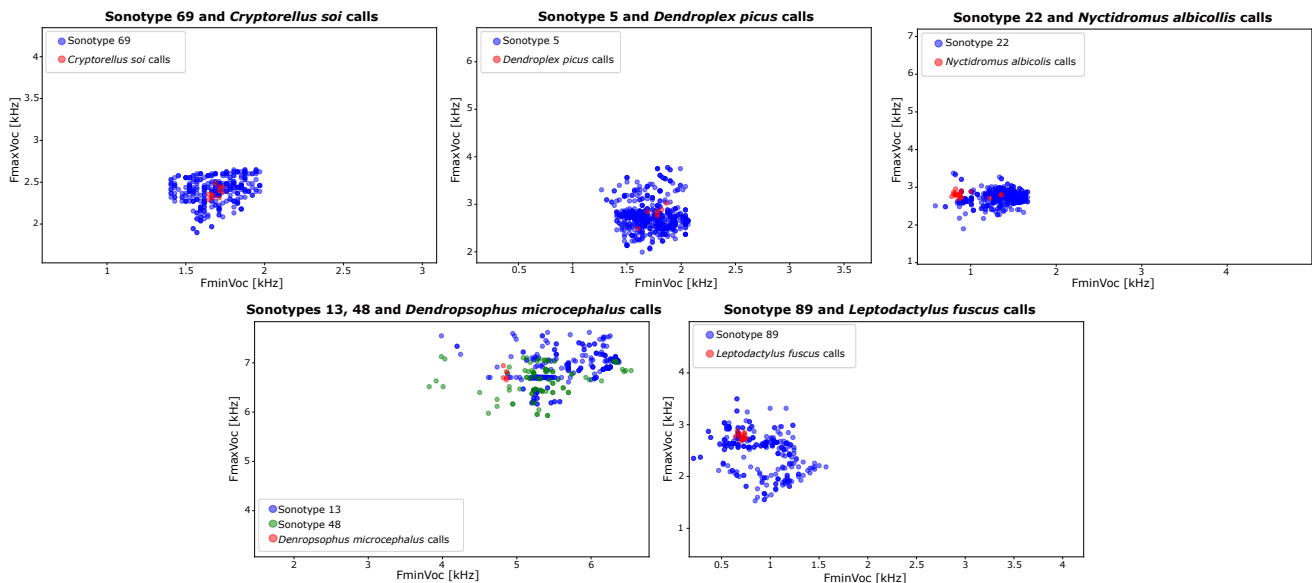

Figure 2. Clustering association results displaying the relationship between sonotypes and species-specific calls. Scatter plots show the minimum frequency (FminVoc) and maximum frequency (FmaxVoc) features for sonotypes (blue dots) and manual call segments (red dots) for select bird and anuran species.

extracted from public datasets, with avian calls being curated from xeno-canto [25] and anuran calls sourced from AmphibiaWeb [26]. This alternative provides valuable support for experts by minimizing the requirement for extensive labeled data to train models.

Figure 2 presents a visual representation of two pivotal acoustic features for the automatic association of species. Each blue and green dot in this figure represents an acoustic segment automatically generated utilizing Guerrero et al. [10] proposed framework. Conversely, the red dots depict manually extracted segments from public databases, showcasing the intersection and validation of automated clustering against recognized species-specific acoustic signatures.

In the case of anuran species *Dendropsophus microcephalus*, it was observed that more than one cluster could be associated with a single species (sonotype 13 and 48). This represents an advantage, as it allows for the description of the variability in the species' calls across multiple clusters, thereby revealing distinct vocal patterns. Such granularity in clustering could prove beneficial for future analyses, offering insights into the diverse calls produced by species.

Due to the need for more information on species present in the study location in public databases, we were able to automatically associate just 5 of 11 audible species that we know are present on the site. Labels for the other species identification were just used for cluster association validation.

IV. CONCLUSION AND FUTURE WORK

This work provides pivotal insights for researchers seeking to select the most fitting tools for biodiversity exploration via acoustic monitoring and automatic call recognition. Our comparative study assessed the performance of various computational intelligence methods, focusing on both machine and deep learning techniques across different taxonomic groups using audio recordings from an agricultural ecosystem. Evaluating these methods with the F1-Score metric, we uncovered the critical need to adapt the choice of methodology to the specific requirements of targeted species. This adaptation is crucial, as we observed significant variances in effectiveness across birds, amphibians, and mammals, each presenting unique challenges that demand specific analytical solutions. Additionally, our findings highlight the importance of comprehensive input preparation for model training (labels, audio formats, among others) and understanding the specifications and requirements of each tool to maximize the advantages of passive acoustic monitoring and automatic detection in biodiversity studies. We also present an approach to cluster interpretation and species association using minimal data, making clustering a promising alternative for grouping patterns and labeling them without extensive expert intervention when labeled data is scarce.

Our results underscore the importance of selecting methodologies based on the specific biological question, whether it involves identifying calls from a particular species or detecting multiple species within an ecosystem. Supervised learning methods, particularly CNNs, are highly effective when sufficient labeled data is available and when the focus is on specific species identification. In contrast, unsupervised methods provide valuable alternatives in data-scarce environments, especially when the goal is to identify the biophonic components of the landscape. Furthermore, the significance of streamlining parameter configuration and adopting strategies that enhance the reproducibility of results was emphasized. For instance, leveraging Koogu to train classifiers from Raven-exported selection tables facilitates the sharing and replication of experimental data among collaborators, potentially enhancing model accuracy through retraining with Koogu-generated results, in a similar way as the methodology presented by Guerrero et al. [10].

By comparing machine learning and deep learning approaches, we pinpointed machine learning's superior capability in understanding and interpreting the features pivotal to the success of the learning model. This insight was particularly evident in our analysis with the LAMDA 3pi clustering algorithm, where features extracted from species calls, commonly utilized by experts for manual identification, aided significantly in cluster interpretation through membership degree information. This methodology not only deepens our comprehension of clustering results but also illustrates its applicability in other scientific domains facing similar challenges of undefined cluster numbers and a scarcity of labeled examples for supervised learning.

While we successfully established automatic species-to-cluster associations based on frequency attributes such as minimum and maximum frequencies, bandwidth, and peak frequency, the overlap of these features among different species poses a challenge in discrimination. To overcome this, employing additional information, such as unique signal features identifying vocalization differences within the same frequency ranges, may prove effective. This approach can significantly improve species differentiation, showing the importance of selecting and fine-tuning features critical to accurately identifying and conserving biodiversity.

Future research should focus on advancing supervised methods to reduce dependence on large amounts of labeled data by using novel computational frameworks, such as few-shot learning. Additionally, addressing the challenge of model parameter setting is crucial in practical applications. From an unsupervised perspective and related to clustering interpretation, future research should explore the clustering variability observed within species calls, where multiple clusters can be associated with a single species. This approach allows for the analysis of call patterns, their features, and the specific circumstances under which they are produced, providing a deeper understanding of species dynamics. Integrating acoustic data with other ecological datasets, such as habitat characteristics and climate data, can provide a deeper understanding of ecosystem health and dynamics, leading to more robust tools for biodiversity monitoring and conservation

REFERENCES

[1] J. Gibbs, H. Snell, and C. Causton, "Effective monitoring for adaptive wildlife management: Lessons from the galapagos islands," *Journal of Wildlife Management*, vol. 63, pp. 1055–1065, 1999. DOI: 10.2307/3802825.

[2] B. Pijanowski *et al.*, "Soundscape ecology: The science of sound in the landscape," *BioScience*, vol. 61, pp. 203–216, 2011. DOI: 10.1525/bio.2011.61.3.6.

[3] P. Wrege, E. Rowland, S. Keen, and Y. Shiu, "Acoustic monitoring for conservation in tropical forests: Examples from forest elephants," *Methods in Ecology and Evolution*, vol. 8, 2017. DOI: 10.1111/2041-210X.12730.

[4] M. Acevedo and L. Villanueva-Rivera, "Using automated digital recording systems as effective tools for the monitoring of birds and amphibians," *Wildlife Society Bulletin*, vol. 34, pp. 211–214, 2006. DOI: 10.2193/0091-7648(2006)34[211:UADRSA]2.0.CO;2.

[5] R. Gibb, E. Browning, P. Glover-Kapfer, and K. E. Jones, "Emerging opportunities and challenges for passive acoustics in ecological assessment and monitoring," *Methods in Ecology and Evolution*, vol. 2019, no. September 2018, pp. 169–185, 2018, ISSN: 2041210X. DOI: 10.1111/2041-210X.13101.

[6] D. Stowell and J. Sueur, "Ecoacoustics: Acoustic sensing for biodiversity monitoring at scale," *Remote Sensing in Ecology and Conservation*, vol. 6, pp. 217–219, 3 2020. DOI: 10.1002/rse2.174.

[7] Z. Zhao *et al.*, "How well do acoustic indices measure biodiversity? computational experiments to determine effect of sound unit shape, vocalization intensity, and frequency of vocalization occurrence on performance of acoustic indices," *Ecological Indicators*, vol. 107, 2019.

[8] J. LeBien *et al.*, "A pipeline for identification of bird and frog species in tropical soundscape recordings using a convolutional neural network," *Ecological Informatics*, vol. 59, p. 101 113, 2020. DOI: 10.1016/j.ecoinf.2020.101113.

[9] J. Xie *et al.*, "Acoustic classification of frog within-species and species-specific calls," *Applied Acoustics*, vol. 131, pp. 79–86, 2018. DOI: 10.1016/j.apacoust.2017.10.024.

[10] M. J. Guerrero, C. L. Bedoya, J. D. López, J. M. Daza, and C. Isaza, "Acoustic animal identification using unsupervised learning.," *Methods in Ecology and Evolution*, vol. 14, pp. 1500–1514, 2023. DOI: 10.1111/2041-210X.14103.

[11] Z. J. Ruff, D. B. Lesmeister, C. L. Appel, and C. M. Sullivan, "Workflow and convolutional neural network for automated identification of animal sounds," *Ecological Indicators*, vol. 124, p. 107 419, 2021. DOI: 10.1016/j.ecolind.2021.107419.

[12] Avisoft Bioacoustics, *Avisoft-saslab pro: Sound analysis and synthesis laboratory*, https://www.avisoft.com/, Accessed: 2024-05-20, 2024.

[13] T. Aide *et al.*, "Real-time bioacoustics monitoring and automated species identification," *PeerJ*, vol. 103, 2013. DOI: doi.org/10.7717/peerj.103.

[14] Wildlife Acoustics, "Kaleidoscope pro 5 user guide," *Wildlife Acoustics, Inc.: Maynard, MA, USA*, 2020.

[15] J. Xie *et al.*, "Frog calling activity detection using lightweight cnn with multi-view spectrogram: A case study on kroombit tinker frog," *Machine Learning with Applications*, vol. 7, p. 100 202, 2021. DOI: 10.1016/j.mlwa.2021.100202.

[16] C. L. Bedoya and L. E. Molles, "Acoustic censusing and individual identification of birds in the wild," *bioRxiv*, 19 2021. DOI: 10.1101/2021.10.29.466450.

[17] C. Corrada Bravo, R. Álvarez Berríos, and T. Aide, "Species-specific audio detection: A comparison of three template-based detection algorithms using random forests," *PeerJ Computer Science*, vol. 113, 2017. DOI: 10.7717/peerj-cs.113.

[18] M. Malfante, M. Dalla Mura, J. Mars, and C. Gervaise, "Automatic fish sounds classification," *Journal of the Acoustical Society of America*, vol. 139, pp. 2115–2116, 2016. DOI: 10.1121/1.4950295.

[19] N. K. Widyastuti, A. Aibinu, M. Salami, R. Akmeliawati, and A. Muthalif, "Animal sound activity detection using multi-class support vector machines," *4th international conference on mechatronics (ICOM). IEEE*, 2011. DOI: 10.1109/ICOM.2011.5937122.

[20] C. Bedoya, C. Isaza, J. M. Daza, and J. D. López, "Automatic recognition of anuran species based on syllable identification," *Ecological Informatics*, vol. 24, pp. 200–209, 2014. DOI: 10.1016/j.ecoinf.2014.08.009.

[21] F. Michaud, J. Sueur, M. Le Cesne, and S. Haupert, "Unsupervised classification to improve the quality of a bird song recording dataset," *Ecological Informatics*, vol. 74, p. 101 952, 2022. DOI: 10.1016/j.ecoinf.2022.101952.

[22] L. S. M. Sugai, T. S. F. Silva, J. W. Ribeiro, and D. Llusia, "Terrestrial passive acoustic monitoring: Review and perspectives," *BioScience*, vol. 69, pp. 5–11, 1 2019, ISSN: 15253244. DOI: 10.1093/biosci/biy147.

[23] B. Miller, S. Madhusudhana, M. Aulich, and N. Kelly, "Deep learning algorithm outperforms experienced human observer at detection of blue whale d-calls: A double-observer analysis," *Remote Sensing in Ecology and Conservation*, vol. 9, 2023. DOI: 10.1002/rse2.297.

[24] Cornell Lab of Ornithology, *Raven pro: Interactive sound analysis software (version 1.6)*, https://ravensoundsoftware.com/, Ithaca, NY: The Cornell Lab of Ornithology. Accessed: 2024-05-20, 2024.

[25] Xeno-canto Foundation, *Xeno-canto: Sharing bird sounds from around the world*, https://www.xeno-canto.org, Accessed: 2024-05-20, 2024.

[26] AmphibiaWeb, *Amphibiaweb: Information on amphib-
ian biology and conservation*, https://amphibiaweb.org,
Accessed: 2024-05-20, 2024.