Overview of BirdCLEF+ 2025: Multi-Taxonomic Sound Identification in the Middle Magdalena, Colombia

Juan Sebastián Cañas[1,2], Stefan Kahl[3,4], Tom Denton[5], Maria Paula Toro-Gómez[2], Susana Rodriguez-Buritica[2], Jose Luis Benavides-Lopez[6], Juan Sebastián Ulloa[2], Paula Caycedo-Rosales[7], Holger Klinck[3], Hervé Goëau[8], Willem-Pier Vellinga[9], Robert Planqué[9] and Alexis Joly[10]

[1]*Centre for Biodiversity and Environment Research, University College London, London WC1E 6BT, UK*

[2]*Instituto de Investigación de Recursos Biológicos Alexander von Humboldt, Bogotá, Colombia*

[3]*K. Lisa Yang Center for Conservation Bioacoustics, Cornell Lab of Ornithology, Cornell University, Ithaca, USA*

[4]*Chemnitz University of Technology, Chemnitz, Germany*

[5]*Google Deepmind, San Francisco, USA*

[6]*Departamento de Ciencias Biológicas, Universidad de los Andes, Bogotá, Colombia*

[7]*Fundación Biodiversa Colombia*

[8]*CIRAD, UMR AMAP, Montpellier, France*

[9]*Xeno-canto Foundation, Groningen, Netherlands*

[10]*Inria, LIRMM, University of Montpellier, CNRS, Montpellier, France*

Abstract

The BirdCLEF+ 2025 challenge focused on the simultaneous acoustic identification of birds, amphibians, mammals and insects in the Middle Magdalena Valley, a biodiversity hotspot in Colombia. This edition aimed to advance passive acoustic monitoring by tasking participants with developing reliable systems for detecting and identifying multi-taxonomic vocalizations from extensive soundscape recordings. Using training data provided by museum collections, citizen science projects and new unlabeled soundscapes, participants addressed the challenge of out-of-distribution generalization under field conditions and limited training data for many species. Participants used data augmentation, pseudo-labeling, and self-training to enhance model robustness and accuracy, often refining pseudo-labels iteratively. For improved scores and runtime efficiency, teams commonly employed Test-Time Augmentation, ensemble methods, and optimized inference with dominant Sound Event Detection and CNN-based models, frequently pretraining on external datasets. The highest-scoring submission achieved an ROC-AUC score of 0.930 on the private leaderboard (0.933 on the public leaderboard), with the top 10 systems differing by only 0.9% in their scores.

Keywords

LifeCLEF, insect, amphibian, mammal, bird, song, call, species, retrieval, audio, collection, identification, fine-grained classification, evaluation, benchmark, bioacoustics, passive acoustic monitoring, PAM

1. Introduction

Some of the world's most biodiversity-rich regions are also those where socioeconomic conflicts run deepest [1]. These areas often lack robust environmental governance, which heightens the tension between conservation and economic exploitation. This institutional fragility exacerbates pressures on ecosystems, undermining both ecological integrity and community well-being. One such region is the Middle Magdalena Valley in Colombia, one of the world's most biodiverse areas, yet it is undergoing rapid land-use intensification [2].

The Middle Magdalena Valley is a vital habitat for numerous taxonomic groups, including mammals, amphibians, birds, and insects [3, 4, 5, 6, 7, 8], thriving in remarkable ecosystems such as humid tropical lowland forests and extensive wetlands. However, economic development in the region—driven by
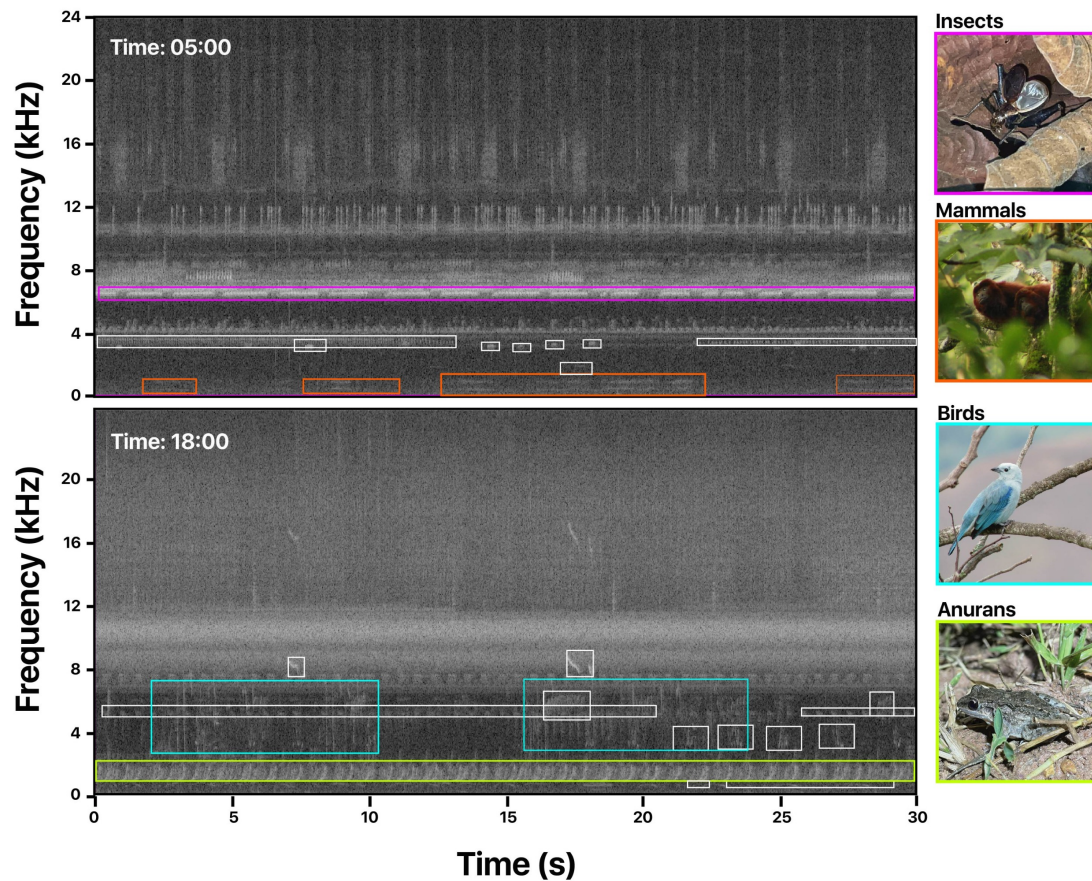
Figure 1: Soundscapes are multi-taxonomic. The dawn soundscape at the top is composed of insects of the Gryllidae family above 4 kHz and primates of the *Allouatta seniculus* species in the lowest part. Between these two signals, additional insect sounds reveal the presence of various individuals both within and outside the same taxonomic group. The bottom spectrogram is a representative soundscape of the dusk composed simultaneously by anurans (*Leptodactylus fragilis*), birds (*Thraupis episcopus*), and multiple overlapping insect signals.

cattle ranching, mineral extraction, and oil palm cultivation—is severely impacting biodiversity and diminishing Nature's Contributions to People (NCP) [9, 10], including water quality and regulation, soil fertility, and carbon sequestration [11, 12, 13, 14]. Therefore, it is essential to design and deploy practical biodiversity diagnostic tools to assess environmental dynamics. In doing so, the community and decision-makers will be better equipped to implement informed, timely strategies that harmonize human development with ecosystem resilience.

In this context, robust biodiversity monitoring is fundamental for rapidly and effectively assessing ecosystem health. For instance, precisely measuring the impact of restoration activities is crucial for identifying optimal treatments that lead to desired ecological outcomes [15, 16]. Acoustic data emerges as a powerful ecological signal for this purpose. Specifically, passive acoustic monitoring (PAM) [17, 18], combined with deep learning models [19] for data analysis, offers a promising approach to inform the efficacy of interventions and track long-term ecological changes. While several studies have explored using sound to evaluate restoration, these approaches primarily rely on the presence of one taxonomic group, such as birds and insects [20, 21], as a proxy for overall diversity.

However, studying entanglement patterns between taxonomic groups could significantly advance our understanding of complex ecological processes, as some studies have shown when examining patterns of presence and absence of different taxonomic groups in the tropical forest [22]. A crucial step for generating time series of species presence and absence required for such analysis is the construction of

highly curated datasets. These datasets are essential for training and testing deep learning models before their broad application to PAM data. Previous works have curated datasets for birds [23], insects [24], amphibians [25], and mammals [26]. While recent work has been merging different sources of data to analyze multi-taxonomic approaches in bioacoustics [27, 28], none have yet considered the co-existence of multiple taxonomic groups in the same soundscape, which are especially rich in the Neotropics (Figure 1), where co-ocurrence, overlapping, and different levels of activity are present in acoustic space [29].

Despite previous works in datasets and automatic species identifiers, there are no PAM datasets that consider the ubiquitous multi-taxonomy of the soundscapes. Furthermore, there are no multi-taxonomic automatic models, as in the case of MegaDetector in camera trapping [30], that can be used as a backbone for different applications in PAM. To address both challenges, we present:

- The **ESMT** (El Silencio Multi-Taxonomic) dataset, composed of two parts: 1) 770 strongly-labeled soundscapes representing 15k bounding boxes of 4 taxonomic groups simultaneously singing in the Middle Magdalena Valley and 2) 11340 unlabeled soundscapes in the same region.

- The **BirdCLEF+ 2025** (Bird Recognition Challenge), an integral part of LifeCLEF 2025 [31], tasked participants with identifying bird, mammal, insect, and amphibian calls within soundscapes from the Middle Magdalena Valley. The competition ended with 9,829 registrations and 2,757 participants on 2,161 teams. We had 76,381 submissions from 86 countries.

2. El Silencio Multi-Taxonomic Dataset

In this section, we describe the construction of the dataset that we release in this work: the **ESMT** dataset. First, we selected a dedicated subset for strong-labelling annotation. Next, we created bounding boxes over frequently observed sonotypes across various taxa. Finally, we assign a taxonomic identification to the sonotypes. In addition, we also selected unlabeled soundscapes for further exploration. The dataset is made publicly available[1].

2.1. Data collection

We deployed recorders across the Middle Magdalena River Valley, around the forests of the Barbacoas wetlands (Figure 2). We used a stratified sampling design across properties and areas with contrasting compositions of forest and pasture. We deployed Audiomoth v1.2.0 [32] passive recorders during March and August 2023. Recorders were located 1.5m from the ground, programmed to capture one minute of sound every five minutes with a sampling rate of 48 kHz.

2.2. Labeled soundscapes

Selection: Seven sites were selected with different forest compositions. A random subset of recordings within the 5-7 AM and 5-7 PM time frames was selected for annotation (110 per site). This resulted in a total of 770 recordings, amounting to 12.8 hours

Annotation: We randomly selected 10 files of each site to hear common sonotypes. We focus on the most frequent and stereotypical sonotypes to decrease workload. Two expert annotators were in charge of creating strong labels over the entire soundscape. One expert checked all the files searching for birds and mammals (PCR), and the other expert annotated insects and amphibians (MPTG).

Taxonomic identification: Birds and mammals were easily recognized through previous works, expertise, audio repositories and a species list provided by the System of Biodiversity from Colombia (SiB Colombia). Amphibians were identified using a similar route but with some additional confirmation

---

[1]https://github.com/redecoacustica/elsilencio-dataset/
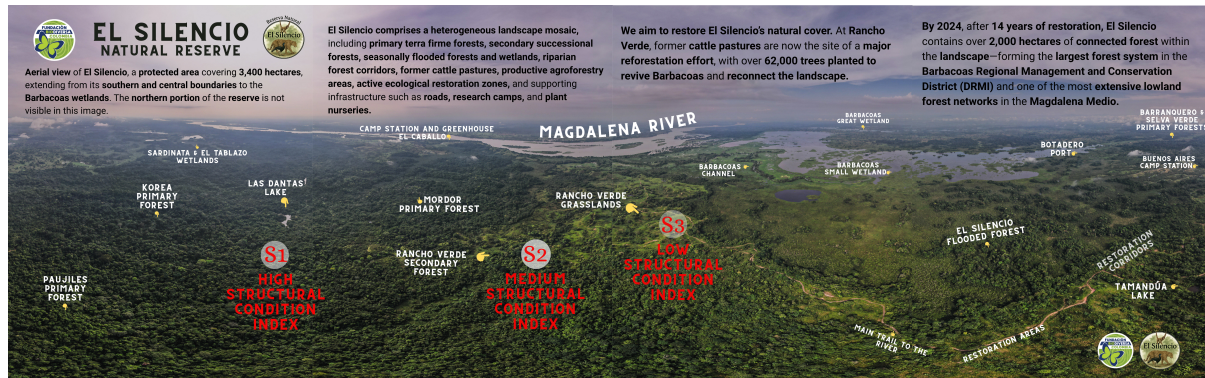
Figure 2: The audio data for this competition was collected from El Silencio Natural Reserve and Research Station in Colombia's Middle Magdalena Valley. AudioMoth ARUs were deployed across a forest regeneration gradient within the reserve to gather the dataset. The recording corresponds to 3 sampling areas with different forest compositions. Area S1 had the highest proportion of forest with high Structural Condition Index. Area S3 has the highest proportion of grasslands and non-forested vegetation (1-3 SCI values), and area S2 is intermediate between the two. (Photo: Santiago Rosado)

between herpetologists. However, the hardest identification process was for insects. After an iterative process [33] that included field work in the reserve and intensive manual verification led by an expert entomologist (JLBL) with the Collection of Environmental Sounds (CSA) in Colombia [34], we identified a subset of the insect sonotypes at the family level. Infrequent sonotypes were not identified.

2.3. Unlabeled soundscapes

From the final 534,420 audio files collected, we randomly selected 11,340 unlabeled soundscapes. We chose that specific quantity to keep the total size of the dataset below 50GB. These files correspond to 63 sites (180 files per site) during all possible hours and days of the collection. We open unlabeled soundscapes to explore potential algorithmic approaches that use unlabeled data to improve species identification models.

3. BirdCLEF+ 2025 Competition Overview

Mobile and habitat-diverse species serve as valuable indicators of biodiversity change, as shifts in their assemblages and population dynamics can signal the success or failure of ecological restoration efforts. These species often respond rapidly to environmental changes, making them particularly useful for detecting early signs of ecological improvement or degradation. However, traditional observer-based biodiversity surveys across large areas are both costly and logistically demanding, often requiring extensive fieldwork, expertise, and repeated visits to remote locations, challenges that limit the frequency and scale of monitoring. In contrast, passive acoustic monitoring (PAM), combined with AI, offers a scalable and non-invasive solution that enables conservationists to collect and analyze vast amounts of ecological data with minimal human presence. PAM systems can operate continuously over extended periods and in challenging environments, capturing the vocal activity of a wide range of taxa, including birds, amphibians, and insects. When paired with automated species identification, it enables researchers to monitor biodiversity across broad spatial and temporal scales, allowing more timely and data-driven reviews of restoration outcomes.

3.1. Goal/Task

This competition aimed to advance automated species identification in soundscape data from the Middle Magdalena Valley of Colombia, including the El Silencio Natural Reserve. Key objectives include detecting species across diverse taxonomic groups, developing machine learning models capable of

recognizing rare and endangered species from limited training data, and leveraging unlabeled data to improve detection and classification performance.

3.2. Evaluation protocol

The challenge was hosted on Kaggle, following a similar evaluation setup as in previous years [35], with hidden test data and a code competition format. We used a variant of macro-averaged ROC-AUC as the evaluation metric, excluding classes with no true positive labels, allowing us to assess model performance without relying on confidence threshold tuning and emphasizing species-level rather than segment-level accuracy [36]. Participants were asked to identify species in short, 5-second audio clips extracted from labeled soundscape recordings, a length chosen to balance signal clarity with adequate context. The dataset was kept under 50 GB to ensure accessibility and ease of use. To further support participants, we provided starter code and documentation to help newcomers get started quickly.

3.3. Time limits

Competitors were limited to 90 minutes of inference time on a CPU. This ensures that models are cost-effective for real-world usage. A side effect is reducing the impact of *ensembling*, a common Kaggle tactic obscuring underlying model quality.

3.4. Dataset for the competition

Building on lessons from previous editions, we refined the task to encourage participants to design models tailored to the unique challenges of the competition. Training and test data were carefully selected to reflect a range of bird and non-bird taxa[2], supporting this goal. As in past years, Xeno-canto [37] remained the main source of training data, complemented by expertly annotated soundscape recordings for testing. This year, we expanded the dataset to include contributions from iNaturalist [38] and the Collection of Environmental Sounds (CSA) of the Humboldt Institute [34, 39], with a focus on underrepresented species, those ecologically important but difficult to detect due to rarity or elusive behavior. The training dataset included commonly occurring species identified via eBird and iNaturalist observation data, supporting the development of robust models in cases where the target species composition is unknown. As a result, some species were present in the training data but absent from the test data, while still being representative of the target region.

Test data sources were selected to capture a broad range of acoustic environments, incorporating differences in call density, background noise, and recording formats (mono and stereo). Species labels were excluded when fewer than five training recordings were available or when species identification could not be confirmed with certainty. Unlabeled training data, designed to resemble the test set, were also included to encourage exploration of semi-supervised and self-supervised learning techniques. In total, the dataset consisted of more than 38,000 labeled training recordings covering 206 species, along with 705 one-minute soundscape recordings for testing and evaluation.

4. Results of BirdCLEF+ 2025

A total of 2,025 teams with nearly 2,569 competitors participated in the BidCLEF+ 2025 competition, submitting a total of 70,674 runs. As in recent years, two-thirds of the test data was allocated to the private leaderboard and one-third to the public leaderboard. Based on the ROC-AUC metric, the baseline score was 0.5, with random confidence scores for all birds across all segments. The highest-scoring submission achieved 0.930 (0.933 on the public leaderboard), with the top 10 systems differing by only 0.9% in their scores. The top 25 participant scores were above 0.905 (Figure 3).

---

[2]We therefore renamed the competition from BirdCLEF to BirdCLEF+
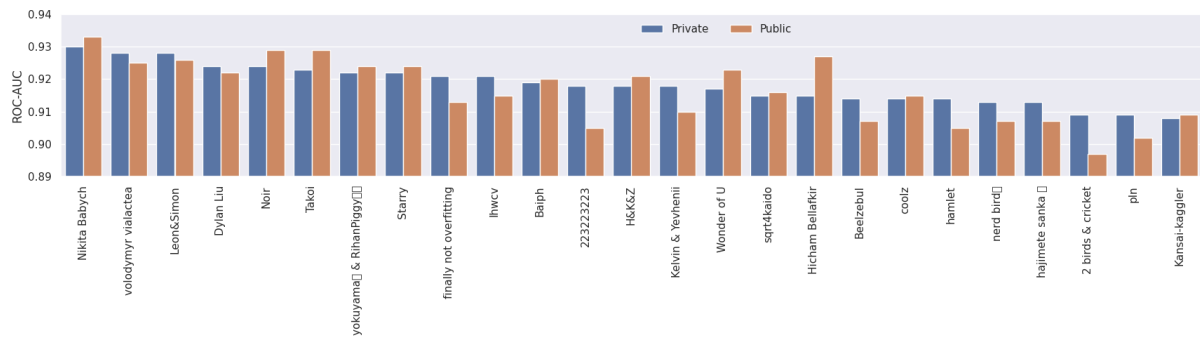
Figure 3: Top 25 private leaderboard scores achieved by the best systems evaluated within the multi-taxonomic identification task of LifeCLEF 2025. Public and private test data were split randomly. The private scores remained hidden until the submission deadline.

The Insecta class presented the most significant challenge in the competition, registering a mean ROC-AUC of 0.667 ± 0.113 across its three considered classes (Figure 4a), which consistently appeared at the lower end of the per-species ranking (Figure 4b). Following this, the Amphibia class achieved a ROC-AUC of 0.840 ± 0.145, notably exhibiting the highest standard deviation, which is also evident in its broad distribution across the per-species ranking. For the dominant Aves class, the mean ROC-AUC was 0.936 ± 0.0809; while some avian species showed minimal performance differences among top participants, others were found lower in the ranking with considerable variation between competitors (Figure 4b). In contrast, the Mammalia class, represented by *Alouatta seniculus*, demonstrated high performance with a ROC-AUC of 0.983 ± 0.020 and a low standard deviation, occupying the upper part of the ranking.

4.1. Online write-up

Across submissions, several common strategies emerged in participants' online write-ups[3]. Data augmentation played a central role, with techniques such as Mixup, Cutmix, Sumix, Frequency and Time Masking, Gain adjustments, Resampling, and FilterAugment widely used. Some teams also introduced external noise, including human speech, to improve model robustness. Undersampled species were typically addressed through upsampling, while pseudo-labeling and self-training on the unlabeled soundscape data proved key for boosting accuracy. These strategies often involved generating pseudo-labels from preliminary models, applying transformations (e.g., power scaling, filtering low-confidence predictions), and iteratively refining the labels. Weighting more confident pseudo-labeled examples more heavily during training also contributed to improved outcomes.

For inference, teams commonly employed Test-Time Augmentation (TTA) by processing overlapping audio segments and smoothing predictions over time, sometimes with delta shifts. Post-processing steps - such as adjusting prediction confidence, applying power-based scaling, or calibrating outputs - were used to further refine model predictions. Ensemble methods, including blending models from different training folds or checkpoints, were instrumental in boosting final scores. To meet runtime constraints, many participants optimized inference speed using tools like ONNX, OpenVINO, and multiprocessing.

The dominant modeling approach was Sound Event Detection (SED), often enhanced with dedicated SED heads. CNN-based models were also widely used, sometimes in hybrid combinations with SED components. EfficientNet backbones were especially popular, though alternatives like RegNet and NFNet also saw successful implementations. Some teams trained separate models for

---

[3]Individual write-ups can be accessed via the "Solution" icon on the leaderboard: https://www.kaggle.com/c/birdclef-2025/leaderboard
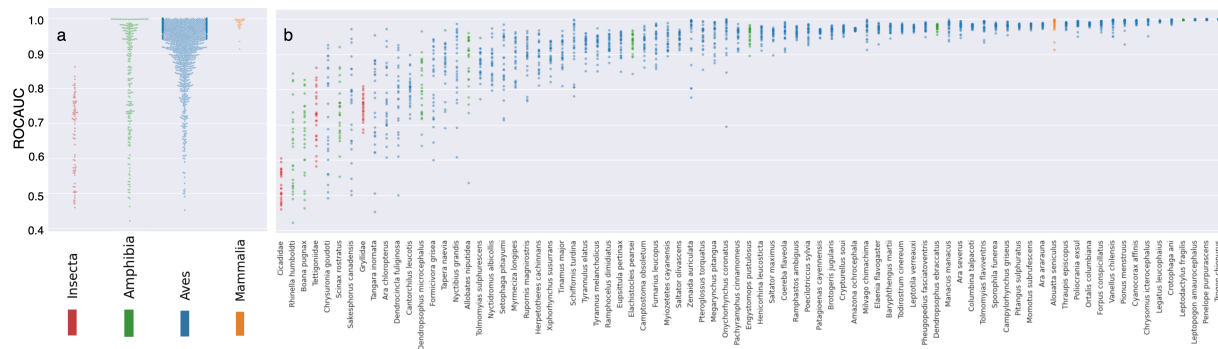
Figure 4: Results per taxonomic group and species for the Top 25 scores. Overall, the hardest group to identify was insects, followed by amphibians, which shows high variability per species. The next group was the Aves group with different results across species. The best class is the Mammalia, with one class in the upper part of the ranking showing high performance and low variability.

taxonomic subgroups (e.g., Amphibia, Insecta), incorporating additional external datasets to improve representation. Input features were typically log-transformed Mel spectrograms, with variation in the number of mel bins, hop sizes, and frequency ranges. A variety of loss functions were explored, including Cross Entropy, BCE With Logits Loss, and Focal Loss variants, with some evidence suggesting Focal or Cross Entropy loss could offer marginal improvements with appropriate tuning. Pretraining model backbones on large external datasets such as Xeno-Canto prior to fine-tuning on the competition data significantly boosted early performance.

4.2. Working notes

We accepted four working notes for the proceedings, which document the approaches and methodologies used by individual teams:

Tan & Wang [40]: The authors developed an end-to-end classification model that uses two parallel input branches (Dual Branch Network) to process Mel-spectrogram and MFCC features, respectively. MFCCs are fed into a ResNet50 pretrained on ImageNet, while Mel features are passed through a randomly initialized ConvNeXt-v2. The feature representations from both branches are fused late in the pipeline to produce final species predictions. The study evaluates different combinations of pretrained and randomly initialized backbones, with a focus on understanding how complementary audio representations and model initialization strategies affect classification performance.

Gokulnath et al. [41]: Adopting a modular approach, this team frames bird species identification as a set of binary classification tasks—one per species. Rather than using a multi-label model, the authors treat the task as 206 independent detection problems, enabling species-specific data augmentation, threshold tuning, and diagnostics. Extensive cross-validation and performance visualizations help analyze which species benefit most from augmentation. The authors argue that this modular design simplifies model interpretation, allows fine-grained tuning, and reduces the complexity of the output layer.

Sydorskyi & Gonçalves [42]: This team employs an ensemble strategy using lightweight CNN architectures—specifically EfficientNetV2-S and NFNet-L0—trained independently on log-Mel spectrograms which were generated from 5-second audio segments. Augmentations such as MixUp and SpecAugment were applied during training. The final predictions are computed by averaging the softmax outputs of 15 different models, leveraging complementary strengths of the individual learners. Ensembling improved prediction accuracy without introducing substantial computational complexity, making it suitable for the competition despite the runtime constraint.

Miyaguchi et al. [43]: This submission presents a token-based classification pipeline that transforms MFCC features into discrete tokens. MFCCs are clustered into 256 discrete tokens using k-means, forming sequences analogous to text. A Word2Vec model is trained on these sequences to learn embeddings, which are then fed into a compact transformer model (the "student") trained to match the outputs of a CNN-based classifier (the "teacher") using KL divergence. This approach results in a model that retains competitive classification performance but is fast enough to process the entire test set in under 5 minutes on CPU.

5. Conclusions and Lessons Learned

The BidCLEF+ 2025 competition showcased remarkable progress in acoustic species identification, drawing 2,025 teams who submitted an impressive 70,674 runs. The top systems achieved exceptional results, with the leading entry hitting a ROC-AUC of 0.930 (0.933 on the public leaderboard) and the top 25 participants consistently scoring above 0.905. This widespread participation and strong performance underscore the significant advancements in bioacoustics species identification.

Participants used several strategies to achieve these results. Key techniques included extensive data augmentation (e.g., Mixup, masking, external noise), upsampling for undersampled species, and crucial pseudo-labeling and self-training on unlabeled data to enhance performance. During inference, Test-Time Augmentation (TTA) and post-processing refined predictions, while ensemble methods further boosted scores. Runtime optimization was also a focus, often through tools like ONNX. The predominant modeling approach was Sound Event Detection (SED), frequently integrated with CNNs (e.g., EfficientNet backbones), with pretraining on large external datasets proving especially effective.

Despite these impressive overall results, a deeper taxonomic analysis revealed persistent challenges. Groups like Insects and Amphibians remain difficult to identify, primarily due to the limited availability of data for these species and taxonomic uncertainty. Furthermore, not all bird species were equally easy to classify, with some showing considerable performance variation among top competitors. Future work should focus on new datasets for these groups and investigate which acoustic characteristics are the strongest determinants of these performance disparities to inform more robust identification models.

Acknowledgments

All results, code notebooks, and forum posts are publicly available at:
https://www.kaggle.com/c/birdclef-2025

Declaration on Generative AI

During the preparation of this work, the author(s) used LanguageTool and Gemini to: Grammar and spelling check. After using these tool(s)/service(s), the author(s) reviewed and edited the content as needed and take(s) full responsibility for the publication's content.

References

[1] B. Vira, A. Kontoleon, Dependence of the poor on biodiversity: which poor, what biodiversity?, Biodiversity conservation and poverty alleviation: Exploring the evidence for a link (2012) 52–84.

[2] G. Forero-Medina, L. Joppa, Representation of global and national conservation priorities by colombia's protected area network, PLoS One 5 (2010) e13210.

[3] F. Vargas-Salinas, A. Aponte-Gutiérrez, Diversidad y recambio de escpecias de anfibios y reptiles entre coberturas vegetales en una localidad del valle del magdalena medio, departamento de antioquia, colombia, Biota colombiana 17 (2016) 117–137.

[4] N. Reyes-Amaya, J. Lozáno-Flórez, D. Flores, S. Solari, Distribution of the spix's disk-winged bat, thyroptera tricolor spix, 1823 (chiroptera: Thyropteridae) in colombia, with first records for the middle magdalena valley, Mastozoología neotropical 23 (2016) 127–137.

[5] W. A. Valencia-Montoya, D. Tuberquia, P. A. Guzmán, J. Cardona-Duque, Pollination of the cycad zamia incognita a. lindstr. & idárraga by pharaxonotha beetles in the magdalena medio valley, colombia: a mutualism dependent on a specific pollinator and its significance for conservation, Arthropod-Plant Interactions 11 (2017) 717–729.

[6] R. Achury, A. Suarez, Richness and composition of ground-dwelling ants in tropical rainforest and surrounding landscapes in the colombian inter-andean valley, Neotropical Entomology 47 (2018) 731–741.

[7] E. Arbeláez-Cortés, D. Villamizar-Escalante, N. Trujillo-Arias, New voucher specimens and tissue samples from an avifaunal survey of the middle magdalena valley of bolívar, colombia, bridge geographical and temporal gaps, The Wilson Journal of Ornithology 132 (2020) 773–779.

[8] H. E. Ramírez-Chaves, et al., Mamíferos de Colombia. v1.14. Sociedad Colombiana de Mastozoología, https://doi.org/10.15472/kl1whs, 2025.

[9] A. Etter, C. McAlpine, H. Possingham, Historical patterns and drivers of landscape change in colombia since 1500: a regionalized spatial approach, Annals of the Association of American Geographers 98 (2008) 2–23.

[10] C. A. C. Ayram, A. Etter, J. Díaz-Timoté, S. R. Buriticá, W. Ramírez, G. Corzo, Spatiotemporal evaluation of the human footprint in colombia: Four decades of anthropic impact in highly biodiverse ecosystems, Ecological Indicators 117 (2020) 106630.

[11] A. Molano, En medio del Magdalena Medio, Centro de Investigación y Educación Popular, 2009.

[12] L. Potter, Colombia's oil palm development in times of war and 'peace': Myths, enablers and the disparate realities of land control, Journal of rural studies 78 (2020) 491–502.

[13] J. Salgado, J. B. Shurin, M. I. Vélez, A. Link, L. Lopera-Congote, C. González-Arango, F. Jaramillo, I. Åhlén, G. De Luna, Causes and consequences of recent degradation of the magdalena river basin, colombia, Limnology and Oceanography Letters 7 (2022) 451–465.

[14] B. Lora-Ariza, A. Piña, L. D. Donado, Assessment of groundwater quality for human consumption and its health risks in the middle magdalena valley, colombia, Scientific Reports 14 (2024) 11346.

[15] T.-A. Natalia, et al., Role of a campesine reserve zone in the magdalena valley (colombia) in the conservation of endangered tropical rainforests, Nature Conservation Research. 8 (2023) 49–63.

[16] P. H. Brancalion, F. Hua, F. H. Joyce, A. Antonelli, K. D. Holl, Moving biodiversity from an afterthought to a key outcome of forest restoration, Nature Reviews Biodiversity (2025) 1–14.

[17] L. S. M. Sugai, T. S. F. Silva, J. W. Ribeiro Jr, D. Llusia, Terrestrial passive acoustic monitoring: review and perspectives, BioScience 69 (2019) 15–25.

[18] R. Gibb, E. Browning, P. Glover-Kapfer, K. E. Jones, Emerging opportunities and challenges for passive acoustics in ecological assessment and monitoring, Methods in Ecology and Evolution 10 (2019) 169–185.

[19] D. Stowell, Computational bioacoustics with deep learning: a review and roadmap, PeerJ 10 (2022) e13152.

[20] J. Müller, O. Mitesser, H. M. Schaefer, S. Seibold, A. Busse, P. Kriegel, D. Rabl, R. Gelis, A. Arteaga, J. Freile, et al., Soundscapes and deep learning enable tracking biodiversity recovery in tropical forests, Nature communications 14 (2023) 6191.

[21] L. A. Do Nascimento, C. Pérez-Granados, J. B. R. Alencar, K. H. Beard, Time and habitat structure shape insect acoustic activity in the amazon, Philosophical Transactions of the Royal Society B 379 (2024) 20230112.

[22] Z. Burivalova, T. Maeda, Y. Rayadin, T. Boucher, P. Choksi, P. Roe, A. Truskinger, E. Game, et al., Loss of temporal structure of tropical soundscapes with intensifying land use in borneo, Science of the Total Environment 852 (2022) 158268.

[23] L. Rauch, R. Schwinger, M. Wirth, R. Heinrich, D. Huseljic, M. Herde, J. Lange, S. Kahl, B. Sick, S. Tomforde, et al., Birdset: A large-scale dataset for audio classification in avian bioacoustics, arXiv preprint arXiv:2403.10380 (2024).

[24] M. Faiß, B. Ghani, D. Stowell, Insectset459: an open dataset of insect sounds for bioacoustic machine learning, arXiv preprint arXiv:2503.15074 (2025).

[25] J. S. Cañas, M. P. Toro-Gómez, L. S. M. Sugai, H. D. Benítez Restrepo, J. Rudas, B. Posso Bautista, L. F. Toledo, S. Dena, A. H. R. Domingos, F. L. de Souza, et al., A dataset for benchmarking neotropical anuran calls identification in passive acoustic monitoring, Scientific Data 10 (2023) 771.

[26] E. Dufourq, I. Durbach, J. P. Hansford, A. Hoepfner, H. Ma, J. V. Bryant, C. S. Stender, W. Li, Z. Liu, Q. Chen, et al., Automated detection of hainan gibbon calls for passive acoustic monitoring, Remote Sensing in Ecology and Conservation 7 (2021) 475–487.

[27] M. Hagiwara, B. Hoffman, J.-Y. Liu, M. Cusimano, F. Effenberger, K. Zacarian, Beans: The benchmark of animal sounds, in: ICASSP 2023-2023 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), IEEE, 2023, pp. 1–5.

[28] M. Chasmai, A. Shepard, S. Maji, G. Van Horn, The inaturalist sounds dataset, Advances in Neural Information Processing Systems 37 (2024) 132524–132544.

[29] L. S. M. Sugai, D. Llusia, T. Siqueira, T. S. Silva, Revisiting the drivers of acoustic similarities in tropical anuran assemblages, Ecology 102 (2021) e03380.

[30] S. Beery, D. Morris, S. Yang, Efficient pipeline for camera trap image review, arXiv preprint arXiv:1907.06772 (2019).

[31] L. Picek, S. Kahl, H. Goëau, L. Adam, et al., Overview of lifeclef 2025: Challenges on species presence prediction and identification, and individual animal identification, in: International Conference of the Cross-Language Evaluation Forum for European Languages, Springer, 2025.

[32] A. P. Hill, P. Prince, E. Piña Covarrubias, C. P. Doncaster, J. L. Snaddon, A. Rogers, Audiomoth: Evaluation of a smart open acoustic device for monitoring biodiversity and the environment, Methods in Ecology and Evolution 9 (2018) 1199–1211.

[33] K. Riede, R. Balakrishnan, Acoustic monitoring for tropical insect conservation, Philosophical Transactions B 380 (2025) 20240046.

[34] A. M. Mendoza-Henao, O. Acevedo-Charry, D. Martínez-Medina, E. Barona-Cortés, S. Córdoba-Córdoba, P. Caycedo-Rosales, J. S. Ulloa, K. G. Borja-Acosta, A. Buitrago-Cardona, H. Pantoja-Sánchez, Past, present, and future of a tropical sounds collection from colombia, Bioacoustics 32 (2023) 474–490.

[35] S. Kahl, T. Denton, H. Klinck, V. Ramesh, V. Joshi, M. Srivathsa, A. Anand, C. Arvind, H. Cp, S. Sawant, et al., Overview of birdclef 2024: Acoustic identification of under-studied bird species in the western ghats, CEUR-WS, 2024.

[36] B. Van Merriënboer, J. Hamer, V. Dumoulin, E. Triantafillou, T. Denton, Birds, bats and beyond: Evaluating generalization in bioacoustics models, Frontiers in Bird Science 3 (2024) 1369756.

[37] Xeno-canto, https://xeno-canto.org/, accessed Feb 13 2025.

[38] iNaturalist, https://www.inaturalist.org/, accessed Feb 13 2025.

[39] Colección de Sonidos Ambientales (CSA) Mauricio Álvarez Rebolledo, https://colecciones.humboldt.org.co/sonidos/, accessed Feb 13 2025.

[40] J. Tan, A. Wang, Dual-branch Network for Species Identification via Passive Acoustic Monitoring, in: CLEF Working Notes 2025, CLEF 2025: Conference and Labs of the Evaluation Forum, September 09–12, 2025, Madrid, Spain, 2025.

[41] S. S. Gokulnath, A. Gaikwad, K. Senthilnathan, C. Das, S. P. Sawant, One Detector per Bird: A Scalable Binary Classification Approach for BirdCLEF 2025, in: CLEF Working Notes 2025, CLEF

2025: Conference and Labs of the Evaluation Forum, September 09–12, 2025, Madrid, Spain, 2025.

[42] V. Sydorskyi, F. Gonçalves, Tackling Domain Shift in Bird Audio Classification via Transfer Learning and Semi-Supervised Distillation: A Case Study on BirdCLEF+ 2025, in: CLEF Working Notes 2025, CLEF 2025: Conference and Labs of the Evaluation Forum, September 09–12, 2025, Madrid, Spain, 2025.

[43] A. Miyaguchi, M. Gustineli, A. Cheung, Distilling Spectrograms into Tokens: Fast and Lightweight Bioacoustic Classification for BirdCLEF+ 2025, in: CLEF Working Notes 2025, CLEF 2025: Conference and Labs of the Evaluation Forum, September 09–12, 2025, Madrid, Spain, 2025.