Beyond Species-Based Assessments: Automatic Acoustic Recognition of Pollinator Function

Alef Iury Siqueira Ferreira
Instituto de Informática
Universidade Federal de Goiás
Goiânia, Goiás, Brazil
alef_iury_c.c@discente.ufg.br

José Neiva Mesquita-Neto
Departamento de Biología y Química
Universidad Católica del Maule
Talca, Maule, Chile
jmesquita@ucm.cl

Nádia Félix Felipe da Silva
Instituto de Informática
Universidade Federal de Goiás
Goiânia, Goiás, Brazil
nadia.felix@ufg.br

Thierson Couto Rosa
Instituto de Informática
Universidade Federal de Goiás
Goiânia, Goiás, Brazil
thierson@ufg.br

Abstract

Although effective crop pollination depends on pollinator function rather than species identity, most monitoring tools rely on labour-intensive taxonomic assessments. Here, we evaluated whether acoustic signals could directly classify the functional roles of flower-visiting bees. Using convolutional neural networks (CNNs) trained on the buzzing sounds of bees visiting blueberry (*Vaccinium corymbosum*) flowers in southern Chile, we achieved a Macro F1-score of 85.5% in distinguishing true pollinators from non-pollinators—exceeding baselines and taxonomic classification. Our findings demonstrate that bee buzzing can reveal ecological function for the first time, enabling the automated recognition of effective pollinators in crop systems. This approach provides scalable tools for pollination management and opens a new direction for ecological monitoring that extends beyond taxonomy to functional roles.

1 Introduction

Deep learning has revolutionized artificial intelligence (AI), enabling sophisticated models for various applications, including bioacoustics (1). However, one might ask whether such models have any practical ecological value. Several AI-based recognition systems have been designed to identify living things based on using species-based assessments of sounds, images and/or videos (e.g., (2; 3; 4; 5; 6)). Although classifications based on classical taxonomy are essential for scientific communication, a functional classification system that takes into account the ecological roles of species can provide a clearer picture of the science of ecosystems (7). Functional roles, which are defined by the ecological services or interactions that species provide, are often assumed based on taxonomic relatedness. However, this assumption can be misleading. In ecology, species are often grouped into functional groups based on their roles in ecosystems, which influence ecosystem services.

Pollination services provide a clear example of how functional grouping can be more relevant than taxonomic identity. For instance, although bees frequently visit flowers, their contributions to blueberry pollination can vary greatly and may even be neutral or negative (8). Functional traits, such as morphology, phenology, and foraging behavior, directly influence pollination efficiency and

39th Conference on Neural Information Processing Systems (NeurIPS 2025) Workshop: AI for Non-Human Animal Communication.

the delivery of pollination services (9; 10; 11). Grouping pollinators by these traits enables the development of more effective strategies to enhance pollination in agricultural systems (12).

Cultivated highbush blueberries (*Vaccinium corymbosum* L.) depend on insect pollinators for optimal fruit production (13; 14; 15; 8) . Effective pollination increases yield, improves fruit set and enhances fruit size (15; 8). However, the efficiency of pollinators varies widely among species that visit the flowers. Blueberry flowers are adapted to buzz pollination, whereby bees vibrate the flowers to release the pollen, producing a characteristic buzzing sound in the process (16; 17). Even among buzz-pollinating bees, the quality of pollen delivery can differ substantially (8). Consequently, local bee communities differ in their effectiveness at pollination, and the presence of inefficient or unsuitable species can reduce pollination services (18; 19; 20).

Recently, convolutional neural networks (CNNs) have been shown to outperform classical machine learning models in the acoustic recognition of highbush blueberry pollinating bees (21). Deep learning approaches, using multi-layered artificial neural networks, proved especially effective in distinguishing among blueberry flower-visiting bee species. These models now represent the state of the art for taxonomic identification of bee species based on their buzzing sounds (21). However, their potential to discriminate true mutualistic pollinators from floral resource thieves remains uncertain.

Given the agricultural importance of blueberries and the need to distinguish visitors based on their functional roles rather than their taxonomy, we aimed to evaluate whether CNN models could differentiate between effective pollinators and non-pollinating visitors of highbush blueberries based on acoustic data. We hypothesized that models excelling in species-level recognition would also perform well in automated functional group recognition. To test this hypothesis, we examined a bee community in southern Chile whose floral visitors had previously been classified as either pollinators or non-pollinators of blueberries (8), and for which acoustic signatures were available. Functional-group-based acoustic recognition has strong potential for practical applications and could open new avenues in computational bioacoustics research.

2 Methodology

We used the dataset of (21), who also trained different deep learning (DL) models for the acoustic detection of bee species visiting blueberry flowers in Chile. Unlike our study, which aimed to recognize functional groups of bees, their analyses focused on recognizing bee species taxonomically. The dataset comprises 518 audio samples, totaling 3,595 buzzing-sound segments (1,728 of which are sonication events), from 15 bee species visiting highbush blueberry (*Vaccinium corymbosum*) flowers in five orchards in southern Chile (Maule and Los Ríos regions) from September to November of 2020 and 2021. Audio sample durations range from five seconds to over one minute and were recorded at a sampling rate of 44.1 kHz. The number of samples per bee species varied significantly, ranging from eight to 108 samples.

2.1 Functional group determination

Blueberry flower visitors were categorised as either true or ineffective pollinators based on the efficiency of conspecific, single-visit pollen deposition. This method provides a direct measure of pollinator performance (22). We adopted (8) classification, which previously evaluated the efficiency of highbush blueberry flower visitors based on conspecific pollen deposition on the stigma using the single-visit test method. Those that deposited significantly more pollen than unvisited flowers were classified as true pollinators, while those that deposited equal to or less than the control of unvisited flowers were classified as ineffective pollinators. We then assigned each recorded bee buzz to one of two functional groups (true or ineffective pollinator) based on the taxon of the visitor.

2.2 Acoustic preprocessing

The acoustic data used was sourced from the pre-existing, curated dataset (see (21)). As this dataset was already preprocessed, no additional modifications were performed. The original authors manually segmented field recordings to isolate bee buzzing sounds, which were subsequently labelled into two distinct categories: sonication and flight. Sonication includes floral buzzing sounds produced by bees vibrating blueberry flowers, while flight encompasses wingbeat sounds from bees flying between flowers. Our experiments incorporated audio segments from both categories. This inclusive approach served two key purposes: it expanded the size of our dataset and ensured the representation of bee species incapable of sonication (23). Furthermore, in line with developing models robust to real-world conditions, no noise removal techniques were applied to the audio signals. This strategy promotes generalisation by training and testing on data that reflects the inherent environmental noise of field recordings (23).

2.3 Data splitting and Data augmentation

The dataset was first partitioned into training, validation, and test subsets. To prevent distributional bias, we employed a stratified splitting strategy based on bee species. This approach ensures that the proportional representation of each species, and by extension, each functional group, was maintained across all splits. To improve model generalisation and address the challenges posed by a small, imbalanced datasets, we applied a suite of data augmentation techniques during training, with each augmentation having a 50% probability of being applied to a given sample. We employed SpecAugment (24), Random Truncation (RT) and Mixup (25), whose efficacy has been demonstrated in prior work in acoustic domains (26; 27), and on bee acoustics (21; 28). SpecAugment, was used to create more robust features by masking random blocks of an audio's spectrogram; we configured it to apply up to two time masks (maximum 64 frames long) and two frequency masks (maximum 8 bins wide). Additionally, Random Truncation (RT) generated new training instances by randomly extracting and concatenating variable-length segments from the original audio recordings, with segment durations of up to 10 seconds. Finally, Mixup generates synthetic samples by creating a weighted linear combination of two different audio samples and their corresponding labels. It blends different samples using the specified mixing parameter λ drawn from a Beta(α, α) distribution. In our experiments, we used α (alpha) = 0.5 for the Beta distribution, which we empirically found worked best for the dataset and task used.

2.4 Training Configuration

For the classification task, we employed the CNN14 architecture from the Pre-trained Audio Neural Networks (PANNs) framework (26). This model was chosen for its favourable balance of performance and computational cost, as well as its proven efficacy for acoustic recognition of bee species (21), in the context of transfer learning. The model was trained with a batch size of 48, using the AdamW optimiser ($\beta_1 = 0.9$, $\beta_2 = 0.98$, $\epsilon = 1 \times 10^{-8}$), and norm-based gradient clipping (with a maximum threshold of 10). A cosine learning rate scheduler, featuring a 500-step linear warm-up, adjusted learning rates between 1×10^{-4} and 5×10^{-4}, trained for 100 epochs using 3 different seeds. All experiments were performed on a single RTX 5000 GPU equipped with 16 GB of VRAM, with each run taking approximately 6 hours to complete. The source code is available at the following link: `https://github.com/alefiury/Pollinator-Function-Automatic-Acoustic-Recognition`.

2.5 Evaluation methods and baseline establishment

We evaluate models primarily using macro-averaged F1 (Macro-F1), which balances macro precision (Macro-P) and macro recall (Macro-R), making it particularly useful for our dataset with an imbalanced class distribution. For completeness, we also report accuracy (Acc).

To our knowledge, no prior work or established benchmark exists for *functional-group* recognition of pollinators from acoustics—and species-level studies are not directly comparable due to different label spaces and objectives—we contextualise performance with two simple baselines: (i) A majority-class baseline that always predicts the most frequent class ("effective pollinators"); and (ii) a uniform-random baseline that assigns labels at random, averaged over 1,000 independent runs.

3 Results

We found that PANNs, the best-performing CNN model for recognizing bee species based on their buzzing sounds during visits to blueberry flowers in southern Chile, also performed strongly in functional-group assessment. The Macro F1-score of PANNs reached 85.5%, far exceeding both baselines (majority: 44.7%; random: 46.7%; Table 1).

Per-class performance of PANNs was uneven, ranging from 76% for non-pollinators to 95% for true pollinators (Fig. 1). This suggests that the model is more likely to misclassify non-pollinators as pollinators than the reverse. The lower performance for non-pollinators likely reflects class imbalance, as this group was underrepresented (N = 689 audio segments) compared to true pollinators (N = 2,906).

While bee identity is only indirectly linked to the delivery of pollination services, functional group recognition provides a direct measure. Automating the functional recognition of flower-visiting bees is therefore particularly relevant for crop production, where pollination quality depends on reliable pollinators. In this way, farmers, agronomists, and other practitioners could determine the pollination role of visiting bees without relying on expert taxonomic knowledge. Recognising the contribution of bees to crop income could also motivate farmers to adopt practices that support the most effective pollinators , thereby benefiting the broader bee community and promoting both profitable and sustainable agriculture.

Table 1: Predictive performance of a convolutional neural network model (pre-trained audio neural networks, PANNs) was assessed by assigning bees to one of two groups (pollinators or non-pollinators) based on their buzzing sounds when visiting blueberry flowers in southern Chile. Model performance was evaluated using the average (± standard deviation) of the following metrics across three runs with different seeds: Macro F1, Macro Recall, Macro Precision, and Accuracy. These were then compared with two baseline scenarios: (1) the majority class, where all classes are assigned to the majority class and (2) the random baseline, where labels are randomly assigned to one of the classes. This was averaged over 1,000 independent runs.

Method	Macro F1 (%)	Macro Recall (%)	Macro Precision (%)	Accuracy (%)
PANNs	**85.52** $\pm$ 2.49	**85.26** $\pm$ 1.82	**85.82** $\pm$ 3.22	**91.06** $\pm$ 1.70
Majority Baseline	44.70 $\pm$ 0.00	50.00 $\pm$ 0.00	40.42 $\pm$ 0.00	80.83 $\pm$ 0.00
Random Baseline	44.67 $\pm$ 1.72	49.91 $\pm$ 2.37	49.94 $\pm$ 1.47	49.94 $\pm$ 1.86

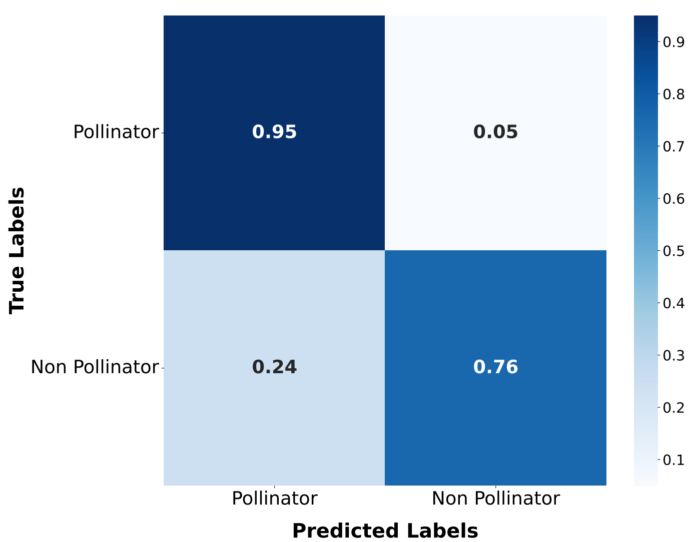

Figure 1: Confusion matrix showing the number of audio segments correctly assigned to each functional group (diagonal elements) versus those misclassified (off-diagonal elements) by the CNN model (Pre-trained Audio Neural Networks, PANNs). The model classified flower-visiting bees into two groups (pollinators or non-pollinators) based on their buzzing sounds when visiting blueberry flowers in southern Chile. Cell colour indicates the number of predicted audio segments, ranging from one (dark blue, all segments correctly predicted) to zero (light blue, no segments predicted).

4 Conclusions

Our results demonstrated that AI-based models developed for species recognition can also be effective in identifying functional groups. Specifically, CNNs that excel at acoustic species recognition can reliably assess pollination function. This indicates that bee buzzing sounds are not only informative for taxonomic classification, but also provide direct access to the ecological roles of species. To our knowledge, this is the first study to achieve functional group recognition of pollinators based on their buzzing sounds. These findings open new research avenues—and potentially a new field—focused on scalable automatic recognition of ecological functions beyond traditional taxonomic frameworks, with potential benefits for low-resource monitoring and more sustainable, function-oriented management in agroecosystems. However, we also acknowledge risks: classification errors could misguide farm decisions if predictions are used without uncertainty estimates and validation; models trained in specific regions, crops, or species pools may not generalize elsewhere; and poorly planned recording campaigns could disturb habitats or enable inappropriate wildlife monitoring in sensitive areas. To mitigate these risks, we recommend calibrated probability outputs and abstention thresholds, cross-site/domain-shift evaluation prior to deployment, adherence to permitting and conservation policies, and human-in-the-loop use so that the system serves as decision support rather than a replacement for expert ecological assessment.

Acknowledgments and Disclosure of Funding

This research was supported by the project "Research and Development of Digital Agents Capable of Planning, Acting, Cooperating, and Learning," funded by the Advanced Knowledge Center in Immersive Technologies (AKCIT). This initiative operates under the Priority Program for the Internet of Things (PPI IoT) of the Ministry of Science, Technology and Innovation (MCTI), with coordination by EMBRAPII. This work was also funded by ANID/CONICYT FONDECYT Regular under grant No. 1231212. The authors also wish to acknowledge the Artificial Intelligence Lab at Recod.ai and the Institute of Computing, University of Campinas, for granting access to the computational infrastructure required for the experiments.

References

[1] Dan Stowell, “Computational bioacoustics with deep learning: a review and roadmap,” *PeerJ*, vol. 10, pp. e13152, 2022.

[2] Jana Wäldchen and Patrick Mäder, “Machine learning for image based species identification,” *Methods in Ecology and Evolution*, vol. 9, no. 11, pp. 2216–2225, 2018.

[3] Anirban Jyoti Hati and Rajiv Ranjan Singh, “Artificial intelligence in smart farms: plant phenotyping for species recognition and health condition identification using deep learning,” *AI*, vol. 2, no. 2, pp. 274–289, 2021.

[4] Shaheer Habib, Mubashir Ahmad, Yasin Ul Haq, Rabia Sana, Asia Muneer, Muhammad Waseem, Muhammad Salman Pathan, and Soumyabrata Dev, “Advancing taxonomic classification through deep learning: A robust artificial intelligence framework for species identification using natural images,” *IEEE Access*, 2024.

[5] Nabanita Das, Neelamadhab Padhy, Nilanjan Dey, Hrithik Paul, and Soumalya Chowdhury, “Exploring explainable ai methods for bird sound-based species recognition systems,” *Multimedia Tools and Applications*, vol. 83, no. 24, pp. 64223–64253, 2024.

[6] Ming Zhong, Ruth Taylor, Naomi Bates, Damian Christey, Hari Basnet, Jennifer Flippin, Shane Palkovitz, Rahul Dodhia, and Juan Lavista Ferres, “Acoustic detection of regionally rare bird species through deep convolutional neural networks,” *Ecological Informatics*, vol. 64, pp. 101333, 2021.

[7] LM Bahr Jr, “Functional taxonomy: an immodest proposal,” *Ecological Modelling*, vol. 15, no. 3, pp. 211–233, 1982.

[8] Benito Cortés-Rivas, Cecilia Smith-Ramirez, Víctor Hugo Monzón, and José Neiva Mesquita-Neto, "Native bees with floral sonication behaviour can achieve high-performance pollination of highbush blueberry in chile," *Agricultural and Forest Entomology*, vol. 25, no. 1, pp. 91–102, 2023.

[9] Patrick Hoehn, Teja Tscharntke, Jason M Tylianakis, and Ingolf Steffan-Dewenter, "Functional group diversity of bee pollinators increases crop yield," *Proceedings of the Royal Society B: Biological Sciences*, vol. 275, no. 1648, pp. 2283–2291, 2008.

[10] Vesna Gagic, Ignasi Bartomeus, Tomas Jonsson, Astrid Taylor, Camilla Winqvist, Christina Fischer, Eleanor M Slade, Ingolf Steffan-Dewenter, Mark Emmerson, Simon G Potts, et al., "Functional identity and diversity of animals predict ecosystem functioning better than species-based indices," *Proceedings of the Royal Society B: Biological Sciences*, vol. 282, no. 1801, pp. 20142620, 2015.

[11] Ben A Woodcock, MPD Garratt, GD Powney, RF Shaw, JL Osborne, J Soroka, SAM Lindström, D Stanley, P Ouvrard, ME Edwards, et al., "Meta-analysis reveals that pollinator functional diversity and abundance enhance crop pollination and yield," *Nature communications*, vol. 10, no. 1, pp. 1481, 2019.

[12] Stephen A Wood, Daniel S Karp, Fabrice DeClerck, Claire Kremen, Shahid Naeem, and Cheryl A Palm, "Functional traits in agriculture: agrobiodiversity and ecosystem services," *Trends in ecology & evolution*, vol. 30, no. 9, pp. 531–539, 2015.

[13] Faye E. Benjamin, James R. Reilly, and Rachael Winfree, "Pollinator body size mediates the scale at which land use drives crop pollination services," *Journal of Applied Ecology*, vol. 51, no. 2, pp. 440–449, 2014.

[14] Jorge B Retamales and James F Hancock, *Blueberries*, vol. 27, Cabi, 2018.

[15] Charles C Nicholson and Taylor H Ricketts, "Wild pollinators improve production, uniformity, and timing of blueberry crops," *Agriculture, Ecosystems & Environment*, vol. 272, pp. 29–37, 2019.

[16] SL Buchmann, "Buzz pollination in angiosperms. in 'handbook of experimental pollination'.(eds ce jones, rj little) pp. 73–113," 1983.

[17] Paul A De Luca, Luc F Bussiere, Daniel Souto-Vilaros, Dave Goulson, Andrew C Mason, and Mario Vallejo-Marín, "Variability in bumblebee pollination buzzes affects the quantity of pollen released from flowers," *Oecologia*, vol. 172, no. 3, pp. 805–816, 2013.

[18] Sarah S Greenleaf and Claire Kremen, "Wild bee species increase tomato production and respond differently to surrounding land use in northern california," *Biological conservation*, vol. 133, no. 1, pp. 81–87, 2006.

[19] O Macias-Macias, J Chuc, P Ancona-Xiu, O Cauich, and JJG Quezada-Euán, "Contribution of native bees and africanized honey bees (hymenoptera: Apoidea) to solanaceae crop pollination in tropical méxico," *Journal of Applied Entomology*, vol. 133, no. 6, pp. 456–465, 2009.

[20] Faye E Benjamin and Rachael Winfree, "Lack of pollinators limits fruit production in commercial blueberry (vaccinium corymbosum)," *Environmental entomology*, vol. 43, no. 6, pp. 1574–1583, 2014.

[21] Alef Iury Siqueira Ferreira, Nádia Felix Felipe Da Silva, Fernanda Neiva Mesquita, Thierson Couto Rosa, Victor Hugo Monzón, and José Neiva Mesquita-Neto, "Automatic acoustic recognition of pollinating bee species can be highly improved by deep learning models accompanied by pre-training and strong data augmentation," *Frontiers in Plant Science*, vol. 14, pp. 1081050, 2023.

[22] Gidi Ne'eman, Andreas Jürgens, Linda Newstrom-Lloyd, Simon G Potts, and Amots Dafni, "A framework for comparing pollinator performance: effectiveness and efficiency," *Biological Reviews*, vol. 85, no. 3, pp. 435–451, 2010.

[23] Alison Pereira Ribeiro, Nádia Felix Felipe da Silva, Fernanda Neiva Mesquita, Priscila de Cássia Souza Araújo, Thierson Couto Rosa, and José Neiva Mesquita-Neto, "Machine learning approach for automatic recognition of tomato-pollinating bees based on their buzzing-sounds," *PLoS computational biology*, vol. 17, no. 9, pp. e1009426, 2021.

[24] Daniel S Park, William Chan, Yu Zhang, Chung-Cheng Chiu, Barret Zoph, Ekin D Cubuk, and Quoc V Le, "Specaugment: A simple data augmentation method for automatic speech recognition," *arXiv preprint arXiv:1904.08779*, 2019.

[25] Hongyi Zhang, Moustapha Cisse, Yann N Dauphin, and David Lopez-Paz, "mixup: Beyond empirical risk minimization," *arXiv preprint arXiv:1710.09412*, 2017.

[26] Qiuqiang Kong, Yin Cao, Turab Iqbal, Yuxuan Wang, Wenwu Wang, and Mark D. Plumbley, "Panns: Large-scale pretrained audio neural networks for audio pattern recognition," *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, vol. 28, pp. 2880–2894, 2020.

[27] Yuan Gong, Yu An Chung, and James Glass, "Psla: Improving audio tagging with pretraining, sampling, labeling, and aggregation," *IEEE/ACM Trans. Audio, Speech and Lang. Proc.*, vol. 29, pp. 3292–3306, jan 2021.

[28] Alef Iury Siqueira Ferreira, Nádia Felix Felipe da Silva, Fernanda Neiva Mesquita, Thierson Couto Rosa, Stephen L Buchmann, and José Neiva Mesquita-Neto, "Transformer models improve the acoustic recognition of buzz-pollinating bee species," *Ecological Informatics*, vol. 86, pp. 103010, 2025.

NeurIPS Paper Checklist

1. **Claims**

   Question: Do the main claims made in the abstract and introduction accurately reflect the paper's contributions and scope?

   Answer: [Yes]

   Justification: The abstract and introduction accurately state the paper's contribution and scope: demonstrating the feasibility of classifying pollinator bee functional groups using deep learning. These claims are substantiated by the methodology (Section 2) and experimental results (Section 3).

   Guidelines:

   - The answer NA means that the abstract and introduction do not include the claims made in the paper.
   - The abstract and/or introduction should clearly state the claims made, including the contributions made in the paper and important assumptions and limitations. A No or NA answer to this question will not be perceived well by the reviewers.
   - The claims made should match theoretical and experimental results, and reflect how much the results can be expected to generalize to other settings.
   - It is fine to include aspirational goals as motivation as long as it is clear that these goals are not attained by the paper.

2. **Limitations**

   Question: Does the paper discuss the limitations of the work performed by the authors?

   Answer: [Yes]

   Justification: As discussed in Section 3, the lower performance on the "non-pollinator" class is likely attributable to the significant class imbalance within the dataset.

   Guidelines:

   - The answer NA means that the paper has no limitation while the answer No means that the paper has limitations, but those are not discussed in the paper.
   - The authors are encouraged to create a separate "Limitations" section in their paper.
   - The paper should point out any strong assumptions and how robust the results are to violations of these assumptions (e.g., independence assumptions, noiseless settings, model well-specification, asymptotic approximations only holding locally). The authors should reflect on how these assumptions might be violated in practice and what the implications would be.
   - The authors should reflect on the scope of the claims made, e.g., if the approach was only tested on a few datasets or with a few runs. In general, empirical results often depend on implicit assumptions, which should be articulated.
   - The authors should reflect on the factors that influence the performance of the approach. For example, a facial recognition algorithm may perform poorly when image resolution is low or images are taken in low lighting. Or a speech-to-text system might not be used reliably to provide closed captions for online lectures because it fails to handle technical jargon.
   - The authors should discuss the computational efficiency of the proposed algorithms and how they scale with dataset size.
   - If applicable, the authors should discuss possible limitations of their approach to address problems of privacy and fairness.
   - While the authors might fear that complete honesty about limitations might be used by reviewers as grounds for rejection, a worse outcome might be that reviewers discover limitations that aren't acknowledged in the paper. The authors should use their best judgment and recognize that individual actions in favor of transparency play an important role in developing norms that preserve the integrity of the community. Reviewers will be specifically instructed to not penalize honesty concerning limitations.

3. **Theory assumptions and proofs**

Question: For each theoretical result, does the paper provide the full set of assumptions and a complete (and correct) proof?

Answer: [NA]

Justification: The paper does not include theoretical results.

Guidelines:

- The answer NA means that the paper does not include theoretical results.
- All the theorems, formulas, and proofs in the paper should be numbered and cross-referenced.
- All assumptions should be clearly stated or referenced in the statement of any theorems.
- The proofs can either appear in the main paper or the supplemental material, but if they appear in the supplemental material, the authors are encouraged to provide a short proof sketch to provide intuition.
- Inversely, any informal proof provided in the core of the paper should be complemented by formal proofs provided in appendix or supplemental material.
- Theorems and Lemmas that the proof relies upon should be properly referenced.

4. **Experimental result reproducibility**

Question: Does the paper fully disclose all the information needed to reproduce the main experimental results of the paper to the extent that it affects the main claims and/or conclusions of the paper (regardless of whether the code and data are provided or not)?

Answer: [Yes]

Justification: The work is based on a publicly available dataset. All experimental details, including data splitting procedures and hyperparameters, are described in Sections 2 and 3.

Guidelines:

- The answer NA means that the paper does not include experiments.
- If the paper includes experiments, a No answer to this question will not be perceived well by the reviewers: Making the paper reproducible is important, regardless of whether the code and data are provided or not.
- If the contribution is a dataset and/or model, the authors should describe the steps taken to make their results reproducible or verifiable.
- Depending on the contribution, reproducibility can be accomplished in various ways. For example, if the contribution is a novel architecture, describing the architecture fully might suffice, or if the contribution is a specific model and empirical evaluation, it may be necessary to either make it possible for others to replicate the model with the same dataset, or provide access to the model. In general. releasing code and data is often one good way to accomplish this, but reproducibility can also be provided via detailed instructions for how to replicate the results, access to a hosted model (e.g., in the case of a large language model), releasing of a model checkpoint, or other means that are appropriate to the research performed.
- While NeurIPS does not require releasing code, the conference does require all submissions to provide some reasonable avenue for reproducibility, which may depend on the nature of the contribution. For example
  (a) If the contribution is primarily a new algorithm, the paper should make it clear how to reproduce that algorithm.
  (b) If the contribution is primarily a new model architecture, the paper should describe the architecture clearly and fully.
  (c) If the contribution is a new model (e.g., a large language model), then there should either be a way to access this model for reproducing the results or a way to reproduce the model (e.g., with an open-source dataset or instructions for how to construct the dataset).
  (d) We recognize that reproducibility may be tricky in some cases, in which case authors are welcome to describe the particular way they provide for reproducibility. In the case of closed-source models, it may be that access to the model is limited in some way (e.g., to registered users), but it should be possible for other researchers to have some path to reproducing or verifying the results.

5. **Open access to data and code**

   Question: Does the paper provide open access to the data and code, with sufficient instructions to faithfully reproduce the main experimental results, as described in supplemental material?

   Answer: [Yes]

   Justification: The source code is publicly available, with a link provided in Section 2.

   Guidelines:

   - The answer NA means that paper does not include experiments requiring code.
   - Please see the NeurIPS code and data submission guidelines (`https://nips.cc/public/guides/CodeSubmissionPolicy`) for more details.
   - While we encourage the release of code and data, we understand that this might not be possible, so "No" is an acceptable answer. Papers cannot be rejected simply for not including code, unless this is central to the contribution (e.g., for a new open-source benchmark).
   - The instructions should contain the exact command and environment needed to run to reproduce the results. See the NeurIPS code and data submission guidelines (`https://nips.cc/public/guides/CodeSubmissionPolicy`) for more details.
   - The authors should provide instructions on data access and preparation, including how to access the raw data, preprocessed data, intermediate data, and generated data, etc.
   - The authors should provide scripts to reproduce all experimental results for the new proposed method and baselines. If only a subset of experiments are reproducible, they should state which ones are omitted from the script and why.
   - At submission time, to preserve anonymity, the authors should release anonymized versions (if applicable).
   - Providing as much information as possible in supplemental material (appended to the paper) is recommended, but including URLs to data and code is permitted.

6. **Experimental setting/details**

   Question: Does the paper specify all the training and test details (e.g., data splits, hyperparameters, how they were chosen, type of optimizer, etc.) necessary to understand the results?

   Answer: [Yes]

   Justification: All experimental details, including data splitting procedures and hyperparameters choises, are described in Section 2.

   Guidelines:

   - The answer NA means that the paper does not include experiments.
   - The experimental setting should be presented in the core of the paper to a level of detail that is necessary to appreciate the results and make sense of them.
   - The full details can be provided either with the code, in appendix, or as supplemental material.

7. **Experiment statistical significance**

   Question: Does the paper report error bars suitably and correctly defined or other appropriate information about the statistical significance of the experiments?

   Answer: [Yes]

   Justification: We report the mean and standard deviation of macro F1-score, macro recall, macro precision, and accuracy for both the main model (3 runs) and the random baseline (1000 runs). These results are presented in Table 1.

   Guidelines:

   - The answer NA means that the paper does not include experiments.
   - The authors should answer "Yes" if the results are accompanied by error bars, confidence intervals, or statistical significance tests, at least for the experiments that support the main claims of the paper.

- The factors of variability that the error bars are capturing should be clearly stated (for example, train/test split, initialization, random drawing of some parameter, or overall run with given experimental conditions).
- The method for calculating the error bars should be explained (closed form formula, call to a library function, bootstrap, etc.)
- The assumptions made should be given (e.g., Normally distributed errors).
- It should be clear whether the error bar is the standard deviation or the standard error of the mean.
- It is OK to report 1-sigma error bars, but one should state it. The authors should preferably report a 2-sigma error bar than state that they have a 96% CI, if the hypothesis of Normality of errors is not verified.
- For asymmetric distributions, the authors should be careful not to show in tables or figures symmetric error bars that would yield results that are out of range (e.g. negative error rates).
- If error bars are reported in tables or plots, The authors should explain in the text how they were calculated and reference the corresponding figures or tables in the text.

8. **Experiments compute resources**

   Question: For each experiment, does the paper provide sufficient information on the computer resources (type of compute workers, memory, time of execution) needed to reproduce the experiments?

   Answer: [Yes]

   Justification: The computational resources used for our experiments are reported in Section 2.

   Guidelines:

   - The answer NA means that the paper does not include experiments.
   - The paper should indicate the type of compute workers CPU or GPU, internal cluster, or cloud provider, including relevant memory and storage.
   - The paper should provide the amount of compute required for each of the individual experimental runs as well as estimate the total compute.
   - The paper should disclose whether the full research project required more compute than the experiments reported in the paper (e.g., preliminary or failed experiments that didn't make it into the paper).

9. **Code of ethics**

   Question: Does the research conducted in the paper conform, in every respect, with the NeurIPS Code of Ethics `https://neurips.cc/public/EthicsGuidelines`?

   Answer: [Yes]

   Justification: The datasets used in this study, which consist of bioacoustic recordings of flower-visiting bees, are both publicly available and were ethically collected. This research fully adheres to the NeurIPS Code of Ethics.

   Guidelines:

   - The answer NA means that the authors have not reviewed the NeurIPS Code of Ethics.
   - If the authors answer No, they should explain the special circumstances that require a deviation from the Code of Ethics.
   - The authors should make sure to preserve anonymity (e.g., if there is a special consideration due to laws or regulations in their jurisdiction).

10. **Broader impacts**

    Question: Does the paper discuss both potential positive societal impacts and negative societal impacts of the work performed?

    Answer: [Yes]

    Justification: Positive impacts (scalable, function-oriented monitoring and support for sustainable farm management) and potential negative impacts (misclassification-driven decisions, limited generalization, habitat disturbance, and surveillance-oriented misuse), along with mitigations, are discussed in Section 4.

Guidelines:

- The answer NA means that there is no societal impact of the work performed.
- If the authors answer NA or No, they should explain why their work has no societal impact or why the paper does not address societal impact.
- Examples of negative societal impacts include potential malicious or unintended uses (e.g., disinformation, generating fake profiles, surveillance), fairness considerations (e.g., deployment of technologies that could make decisions that unfairly impact specific groups), privacy considerations, and security considerations.
- The conference expects that many papers will be foundational research and not tied to particular applications, let alone deployments. However, if there is a direct path to any negative applications, the authors should point it out. For example, it is legitimate to point out that an improvement in the quality of generative models could be used to generate deepfakes for disinformation. On the other hand, it is not needed to point out that a generic algorithm for optimizing neural networks could enable people to train models that generate Deepfakes faster.
- The authors should consider possible harms that could arise when the technology is being used as intended and functioning correctly, harms that could arise when the technology is being used as intended but gives incorrect results, and harms following from (intentional or unintentional) misuse of the technology.
- If there are negative societal impacts, the authors could also discuss possible mitigation strategies (e.g., gated release of models, providing defenses in addition to attacks, mechanisms for monitoring misuse, mechanisms to monitor how a system learns from feedback over time, improving the efficiency and accessibility of ML).

11. **Safeguards**

    Question: Does the paper describe safeguards that have been put in place for responsible release of data or models that have a high risk for misuse (e.g., pretrained language models, image generators, or scraped datasets)?

    Answer: [NA]

    Justification: This research does not involve high-risk models, such as large-scale generative models, or sensitive scraped datasets. Therefore, safeguards for responsible release are not applicable.

    Guidelines:

    - The answer NA means that the paper poses no such risks.
    - Released models that have a high risk for misuse or dual-use should be released with necessary safeguards to allow for controlled use of the model, for example by requiring that users adhere to usage guidelines or restrictions to access the model or implementing safety filters.
    - Datasets that have been scraped from the Internet could pose safety risks. The authors should describe how they avoided releasing unsafe images.
    - We recognize that providing effective safeguards is challenging, and many papers do not require this, but we encourage authors to take this into account and make a best faith effort.

12. **Licenses for existing assets**

    Question: Are the creators or original owners of assets (e.g., code, data, models), used in the paper, properly credited and are the license and terms of use explicitly mentioned and properly respected?

    Answer: [Yes]

    Justification: The paper properly credits the creators of the original dataset and base model via citations in Section 2. Furthermore, the dataset is used in accordance with its Creative Commons Attribution 4.0 International license.

    Guidelines:

    - The answer NA means that the paper does not use existing assets.
    - The authors should cite the original paper that produced the code package or dataset.

- The authors should state which version of the asset is used and, if possible, include a URL.
- The name of the license (e.g., CC-BY 4.0) should be included for each asset.
- For scraped data from a particular source (e.g., website), the copyright and terms of service of that source should be provided.
- If assets are released, the license, copyright information, and terms of use in the package should be provided. For popular datasets, `paperswithcode.com/datasets` has curated licenses for some datasets. Their licensing guide can help determine the license of a dataset.
- For existing datasets that are re-packaged, both the original license and the license of the derived asset (if it has changed) should be provided.
- If this information is not available online, the authors are encouraged to reach out to the asset's creators.

13. **New assets**

    Question: Are new assets introduced in the paper well documented and is the documentation provided alongside the assets?

    Answer: [Yes]

    Justification: We release a public code repository for functional-group classification (link provided in Section 2.4); combined with Sections 2–2.4, the paper specifies splits, hyperparameters, and procedures needed to use the asset reproducibly.

    Guidelines:

    - The answer NA means that the paper does not release new assets.
    - Researchers should communicate the details of the dataset/code/model as part of their submissions via structured templates. This includes details about training, license, limitations, etc.
    - The paper should discuss whether and how consent was obtained from people whose asset is used.
    - At submission time, remember to anonymize your assets (if applicable). You can either create an anonymized URL or include an anonymized zip file.

14. **Crowdsourcing and research with human subjects**

    Question: For crowdsourcing experiments and research with human subjects, does the paper include the full text of instructions given to participants and screenshots, if applicable, as well as details about compensation (if any)?

    Answer: [NA]

    Justification: No crowdsourcing or human subjects were involved; the study uses field recordings of bees.

    Guidelines:

    - The answer NA means that the paper does not involve crowdsourcing nor research with human subjects.
    - Including this information in the supplemental material is fine, but if the main contribution of the paper involves human subjects, then as much detail as possible should be included in the main paper.
    - According to the NeurIPS Code of Ethics, workers involved in data collection, curation, or other labor should be paid at least the minimum wage in the country of the data collector.

15. **Institutional review board (IRB) approvals or equivalent for research with human subjects**

    Question: Does the paper describe potential risks incurred by study participants, whether such risks were disclosed to the subjects, and whether Institutional Review Board (IRB) approvals (or an equivalent approval/review based on the requirements of your country or institution) were obtained?

    Answer: [NA]

Justification: The research does not involve human subjects.

Guidelines:

- The answer NA means that the paper does not involve crowdsourcing nor research with human subjects.
- Depending on the country in which research is conducted, IRB approval (or equivalent) may be required for any human subjects research. If you obtained IRB approval, you should clearly state this in the paper.
- We recognize that the procedures for this may vary significantly between institutions and locations, and we expect authors to adhere to the NeurIPS Code of Ethics and the guidelines for their institution.
- For initial submissions, do not include any information that would break anonymity (if applicable), such as the institution conducting the review.

16. **Declaration of LLM usage**

Question: Does the paper describe the usage of LLMs if it is an important, original, or non-standard component of the core methods in this research? Note that if the LLM is used only for writing, editing, or formatting purposes and does not impact the core methodology, scientific rigorousness, or originality of the research, declaration is not required.

Answer: [NA]

Justification: The core methodology uses convolutional neural networks for acoustic classification; large language models are not part of the technical approach.

Guidelines:

- The answer NA means that the core method development in this research does not involve LLMs as any important, original, or non-standard components.
- Please refer to our LLM policy (`https://neurips.cc/Conferences/2025/LLM`) for what should or should not be described.