Improved Detection of Bird Vocalisations Using BirdNET Embeddings and Machine Learning

Hakan Dogan, *AI and Data Researcher*

***Abstract*—Automated bird sound recognition has become an essential tool for biodiversity monitoring, enabling large-scale species detection from audio recordings. BirdNET is a well-known deep learning-based algorithm that has been trained using a vast dataset of weakly labeled recordings and demonstrated strong performance in identifying bird species. When applied on a certain case such as a specific species or a geographical location, its performance can be leveraged through fine-tuning or incorporating a posterior classification step.**

In this study, the detection of the Eurasian Woodcock (Scolopax rusticola) is investigated, using BirdNET embeddings as feature representations and training a classifier based on them. A strongly labeled dataset is created by manually annotating 97 recent recordings (2023–2024) from Xeno-canto, extracting 501 positive segments and 2,505 negative segments. BirdNET was then evaluated on this dataset, achieving an average precision of 84%. To enhance the detection accuracy, three machine learning classifiers are trained —Support Vector Machine (SVM), Random Forest, and XGBoost— using BirdNET's embeddings as input features. The results demonstrate a significant improvement in classification performance, with overall average precision scores reaching the values of 99–100%, surpassing BirdNET's baseline performance. These findings suggest that a hybrid deep learning and classical machine learning approach can substantially enhance bird species recognition, particularly for challenging acoustic environments.

This work contributes to advancing bioacoustic classification methodologies by demonstrating how deep learning embeddings can be effectively leveraged with traditional classifiers and strongly labeled data, for improved species detection. Future research may explore the applicability of this approach to other species and recording conditions, further refining the bird sound classification systems.

***Index Terms*— artificial intelligence, bird calls, machine learning, audio processing.**

I. INTRODUCTION

Automated bird species identification has become an essential tool in avian ecology, conservation, and biodiversity monitoring. Traditional methods of bird identification, such as direct field observation and manual audio analysis, are time-consuming and require expert knowledge (Kahl et al., 2021). With advancements in deep learning and bioacoustics, machine learning-based approaches have significantly improved the accuracy and scalability of species detection, especially in passive acoustic monitoring (Stowell et al., 2019). Among these methods, BirdNET, a deep neural network trained for bird sound classification, has gained prominence due to its robustness and extensive training on diverse avian vocalizations (Kahl et al., 2021).

The state of the art deep neural networks for bioacoustics utilize large amounts of data and generally classify the vocalisations of many species simultaneously (Lasseck, 2018, 2023, Kahl et al., 2021, Sprengel et al., 2016, Robinson et al., 2024). The problem is handled in an optimization setting for multiple classes and the objective (loss) function is minimized to meet this target. However, under certain considerations, the performance of the developed model for a specific species may also be of importance, i.e. for monitoring purposes and biodiversity assessment (Kramer et al., 2024, Young et al., 2024). For instance, the population change and the evolutionary and ecological dynamics of the Eurasian Woodcock (*Scolopax rusticola*) were investigated in several studies (Aradis et al. 2019, Heward et al. 2024, Bristow et al. 2022, Tuti et al. 2023, Christensen et al. 2017, Engler et al. 2025, Schaly et al. 2024, Prieto et al. 2019, Sládeček et al. 2023). This indicates that, whilst such deep neural networks deliver very robust feature representations (embeddings) in audio data, their performance for an individual species or for a strongly labelled custom dataset may still be prone to improvements (Ghani et al., 2023, Tolkova et al., 2021, Bayat & Işık, 2020). Moreover, species-specific classification accuracy can be

influenced by environmental noise, overlapping bird calls, and intra-species vocal variations (Grill & Schlüter, 2017, Michaud et al. 2023). To address these limitations, a common approach is to extract deep learning-based feature embeddings from pre-trained models and apply a secondary classifier to refine the classification results (Xie et al., 2020, Williams et al. 2024, Ghani et al. 2023). In this study, we apply this technique to enhance the detection accuracy of the Eurasian Woodcock (*Scolopax rusticola*), a nocturnal bird species with distinct vocal characteristics.

The current method leverages BirdNET embeddings as feature vectors and trains a secondary classifier to distinguish between the target and the non-target acoustic events. We evaluate three well-established classification algorithms: Support Vector Machines (SVM), Random Forest (RF), and eXtreme Gradient Boosting (XGBoost). SVMs have been widely used for bioacoustic classification due to their ability to handle high-dimensional feature spaces effectively (Kershenbaum et al., 2016). Random Forests offer robustness against overfitting and interpretability in classification tasks (Liaw & Wiener, 2002), while XGBoost provides strong predictive performance and scalability in structured learning problems (Chen & Guestrin, 2016).

Several studies have demonstrated the effectiveness of combining deep learning embeddings with secondary classifiers in bird sound recognition. For example, Xie et al. (2020) used convolutional neural network (CNN) embeddings with an SVM classifier to improve bird species detection in urban environments. Similarly, Mac Aodha et al. (2018) enhanced classification performance by using deep representations coupled with Random Forest models for species identification in complex soundscapes. Note that transfer learning can also be applied to other research areas such as marine biodiversity (Williams et al. 2024, Ibrahim et al. 2024) and wildlife monitoring (Kath et al., 2024). The present study contributes to improving species-specific detection accuracy of the Eurasian Woodcock and supports the development of more reliable automated bioacoustic monitoring tools.

In the following sections, we detail the methodology used to extract and process BirdNET embeddings, describe the training of the secondary classifiers, and present an evaluation of the system's effectiveness in strongly labeled acoustic datasets. The proposed method aims to provide ecologists with a more accurate tool for species monitoring and conservation research.

II. DATA LABELING

The reliability of any machine learning model heavily depends on the quality of its training and validation data. In the case of BirdNET, the model is trained majorly on data from Xeno-canto, a publicly available repository of bird sound recordings contributed by ornithologists and citizen scientists worldwide. Since its publication in 2021 (Kahl et al., 2021), BirdNET has been widely used for automatic bird sound classification and has demonstrated strong performance in recognizing various species across different environments. However, as with any deep learning model, its effectiveness is influenced by the characteristics of its training data, such as the temporal and geographical distribution of species recordings (Goëau et al., 2020).

To conduct an independent evaluation of BirdNET and further develop an improved classification approach, we have manually annotated 97 audio recordings of the species Eurasian Woodcock, sourced from Xeno-canto (see Fig. 1). A key aspect of the dataset preparation was the deliberate selection of recent recordings from 2023 and 2024, with the objective of creating a presumably unseen evaluation dataset for the BirdNET. This approach helps to assess BirdNET's performance on a dataset, which might differ from those in its original training set due to strong annotations and various background events.

The BirdNET algorithm processes and evaluates audio files at a sampling rate of 48 kHz and operates on 3-second audio segments as input. To align with this framework, annotations were meticulously created in 3-second intervals, ensuring comprehensive coverage of both high-frequency and low-frequency calls of the Eurasian Woodcock (the so-called grunts and squeaks, respectively). This strict segmentation strategy guarantees that no portion of a call lies outside the annotated intervals, preserving the integrity of the training and evaluation data. Furthermore, this approach ensures that both BirdNET and the post-classifier in this study are tested on strongly labeled data, in contrast to the weakly labeled training data originally used for BirdNET. This distinction is crucial, as strongly labeled datasets provide a more precise evaluation of algorithmic performance and reduce ambiguity in classifier training, leading to more reliable detection outcomes.

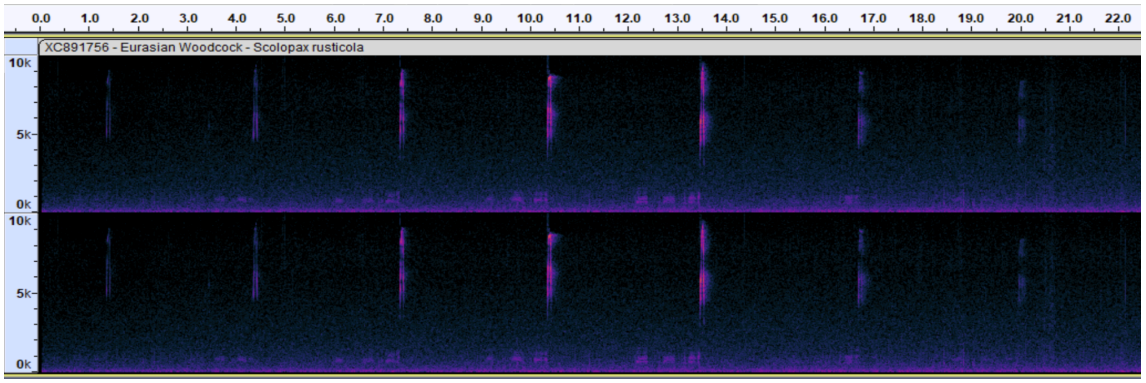

Fig 1. An example spectrogram for an Eurasian Woodcock recording (XC891756). During the flight of the bird, the grunt and the squeak calls as well as the changes to the amplitude strength are visible (created with Audacity).

From these manually annotated recordings, we have extracted a total of 501 individual Eurasian Woodcock (Scolopax rusticola) calls. To create a robust dataset for binary classification, we also included negative segments—audio clips from the same recordings where no Woodcock calls were present. This strategy is essential for ensuring the quantification of false negatives and false positives, similar to the real-world conditions encountered in usual audio recordings (Stowell et al., 2019).

This dataset serves two primary purposes:

1. BirdNET Performance Evaluation – The entire annotated dataset provides a testbed for assessing BirdNET's classification accuracy when applied to new data, allowing to quantify its detection performance under recent real-world conditions.
2. Developing a Secondary Classifier – The extracted embeddings from BirdNET are used to train a separate binary classifier (SVM, Random Forest, XGBoost) that aims to refine and improve detection accuracy, particularly for distinguishing target calls that were initially false negatives.

To ensure transparency and reproducibility, the annotations created in this study are published separately as an independent dataset. The dataset includes a CSV file containing the Xeno-canto file numbers and the timestamps for the positive segments, and is publicly available on Zenodo (Dogan, H., 2025). Since the original audio files are already hosted on Xeno-canto, users can easily retrieve them using the provided file numbers.

III. BirdNET PREDICTIONS

To systematically evaluate BirdNET's performance on the Eurasian Woodcock dataset, we conducted an inference test using the 501 annotated positive segments along with 2,505 negative segments. The inclusion of negative segments is essential for ensuring the robustness of the evaluation metrics, as models trained on imbalanced data can be biased towards the dominant class (Sokolova & Lapalme, 2009).

The results indicate that BirdNET performs quite well, achieving an average precision of 84.3%. This suggests that the model is highly effective in distinguishing Woodcock calls from background noise and other non-target sounds. Furthermore, the precision-versus-threshold and recall-versus-threshold curves provide deeper insights into the model's behavior across different confidence thresholds.

Figure 2 presents the precision vs. threshold curve, which demonstrates a consistent trend of high precision, confirming BirdNET's ability to minimize false positives. For instance, for a threshold value of 0.5, the model gives a precision value of approximately 0.95, indicating the robustness of the algorithm. Figure 3 illustrates the recall vs. threshold curve, where an interesting observation emerges: the recall values increase quite slightly as the threshold value drops in the range between 0.2 and 0.8. This pattern suggests that BirdNET assigns relatively lower probability scores to some weak calls or calls embedded within high levels of background noise. In other words, while the model can detect such calls, it does so with lower confidence, which may impact recall at high threshold values.

BirdNET overall delivers strong performance, characterized by high precision and low false positive rates, making it a reliable baseline for Woodcock detection. These findings provide a solid foundation for developing a post-classification approach, where a separate classifier can further refine the predictions, particularly in borderline cases where BirdNET exhibits lower confidence.

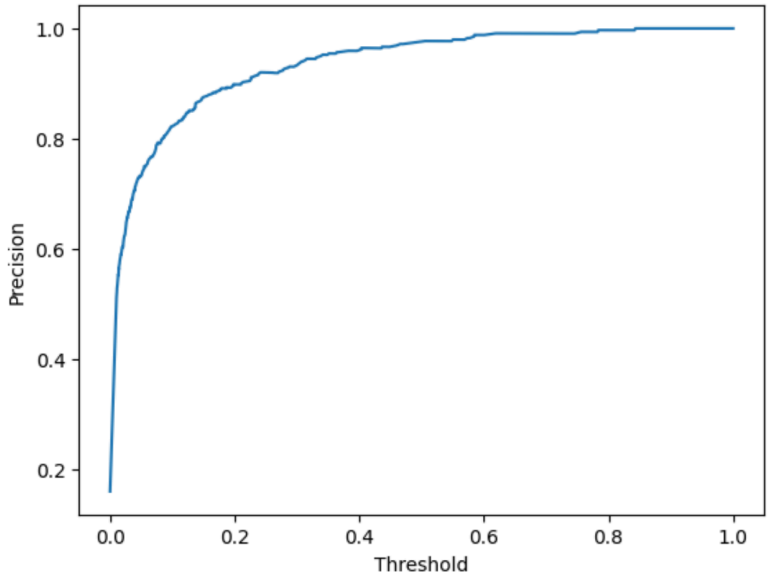

Fig 2. Precision vs. threshold results for BirdNET on the full dataset.

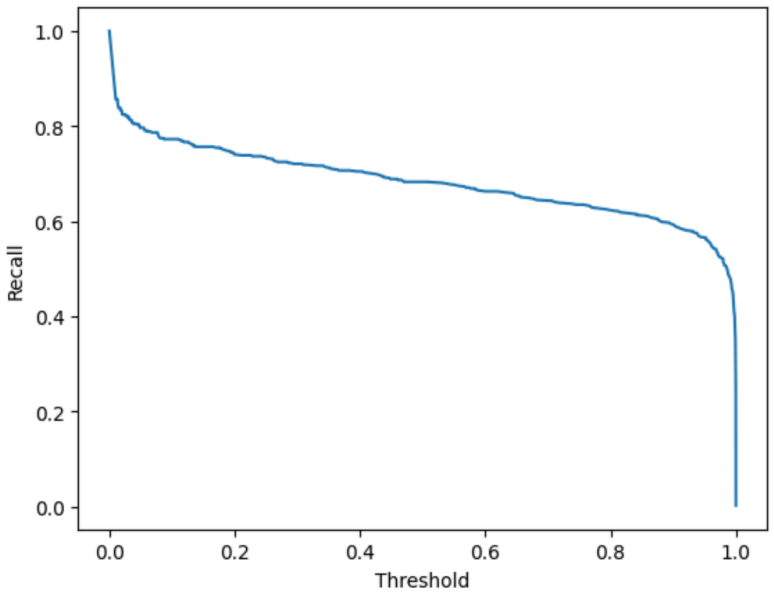

Fig 3. Recall vs. threshold results for BirdNET on the full dataset.

IV. POSTERIOR CLASSIFICATION

A. Cross-Validation Strategy for Training and Testing

To improve the detection accuracy of the Eurasian Woodcock calls, we employ three machine learning classifiers—Support Vector Machine (SVM), Random Forest, and XGBoost—using BirdNET embeddings as input features. A total of 501 positive segments of Woodcock calls were annotated, and to ensure robust model evaluation, a much larger number of 2,505 negative segments was included. This approach addresses two key factors: (i) the diversity of non-target data that may be encountered during real-world applications, and (ii) the possible scarcity of training data for certain species in general.

From the 2,505 negative segments, we created 5 splits as undersampled subsets to establish a balanced dataset between the positive (501) and the negative (501) examples in each fold. Each resulting fold of 1,002 samples (501 positive and 501 negative) is then further split into 5 train-test partitions, allowing for a cross-validation strategy. This procedure helps mitigate any potential overfitting and provides a reliable assessment of each classifier's generalization capability in distinguishing Woodcock calls from non-target sounds in a variety of acoustic contexts.

The final evaluation is presented in the next subsection in a table that displays the cross-validation scores for each classifier across all 5 negative folds. The table has 5 columns corresponding to each of the negative folds and 3 rows representing the three machine learning methods (SVM, Random Forest, and XGBoost). The values in the table will reflect the arithmetic average of the cross-validation results for each classifier on the corresponding negative fold.

B. Performance Evaluation

The mean accuracy values of the trained classifiers are presented in Table 1, whereas Table 2 shows the mean average precision values. Among the classifiers tested, SVM delivered the best performance, achieving an impressive average precision value of 1.0 and a mean accuracy value of 0.997. The high performance of SVM might be attributed to the linear kernel, which works particularly well with the BirdNET embeddings, as these embeddings provide a strong feature representation. This allows SVM to make precise decisions when distinguishing between target calls (Eurasian Woodcock) and non-target events.

TABLE I

MEAN ACCURACY VALUES FOR THREE DIFFERENT MACHINE LEARNING CLASSIFIERS

	Fold 0	**Fold 1**	**Fold 2**	**Fold 3**	**Fold 4**
SVM	0.996	0.997	0.995	0.997	0.998
Random Forest	0.98	0.984	0.977	0.976	0.983
XGBoost	0.978	0.982	0.979	0.982	0.981

Following SVM, the Random Forest classifier achieved an average precision of around 99.9% and a mean accuracy value of 98.1%, showing solid performance, though slightly less optimal than SVM. XGBoost, while still effective, ranked third with an average precision of approximately 99.8% and a mean accuracy value of 98%. Despite being the least performant of the three, XGBoost still demonstrates good discrimination ability and robustness in handling complex data structures.

These preliminary results suggest that SVM is the most effective classifier for this task, likely due to its ability to exploit the linear separability of the feature space created by BirdNET embeddings. However, the overall results suggest that the use of the posterior classifier methodology here, applied on the features from a robust deep learning model, provides a very accurate way of recognizing bird species in audio recordings.

TABLE II

AVERAGE PRECISION VALUES FOR THREE DIFFERENT MACHINE LEARNING CLASSIFIERS

	Fold 0	Fold 1	Fold 2	Fold 3	Fold 4
SVM	1	1	1	1	1
Random Forest	0.999	0.998	0.999	0.999	1
XGBoost	0.998	0.998	0.998	0.998	1

V. CONCLUSIONS

This study demonstrates the effectiveness of combining BirdNET embeddings with classical machine learning classifiers—namely Support Vector Machine (SVM), Random Forest, and XGBoost—to improve the detection of Eurasian Woodcock (Scolopax rusticola) calls. By leveraging the embeddings generated by BirdNET, which was initially trained on weakly labeled data, this research introduces a customized classification approach that enhances the accuracy of bird sound detection, particularly in environments with complex background noise.

The BirdNET algorithm itself achieved an average precision of 84.3% when tested on a dataset of 501 positive segments and 2,505 negative segments, providing a solid baseline for species detection. The precision and recall analyses highlighted the algorithm's stability across various thresholds, although lower confidence was observed for weak calls and those embedded in noisy backgrounds.

When applying the machine learning classifiers to BirdNET embeddings, initial results indicate substantial improvements. The SVM classifier emerged as the most effective, with an average precision of 1.0 in the current case, followed by Random Forest at 99.9% and XGBoost at 99.8%. These findings suggest that SVM, with its linear kernel, is particularly adept at discriminating between target and non-target classes, aided by the strong feature representation provided by BirdNET.

Future work might involve extending the dataset and the methodology to different species and call types. Note that, the strongly labeled dataset created through this work has been made publicly accessible on Zenodo, contributing to the growing body of research data in automated bioacoustic classification.

This approach, combining deep learning-based embeddings with traditional machine learning classifiers, holds great potential for advancing species detection in bioacoustic research and has demonstrated a near-perfect identification of the Eurasian Woodcock species in the current study. It not only provides a robust methodology for a single species but also sets the stage for further exploration into the detection of other species, contributing to the broader field of automated bioacoustic monitoring.

Acknowledgment

The codes used to generate the simulations and the results are available in the following GitHub repository: https://github.com/hdogan84/Woodcock-CNN.

References

Aradis, A., Lo Verde, G., Massa, B. (2019) Importance of millipedes (Diplopoda) in the autumn-winter diet of Scolopax rusticola, *The European Zoological Journal*, doi:10.1080/24750263.2019.1611955, **86**, 1, (452-457).

Bayat S, & Işık G. (2020). Identification of Aras Birds with Convolutional Neural Networks. 4th International Symposium on Multidisciplinary Studies and Innovative Technologies, ISMSIT 2020 - Proceedings.

Blumstein, D.T., Mennill, D.J., Clemins, P., Girod, L., Yao, K., Patricelli, G., Deppe, J.L., Krakauer, A.H., Clark, C., Cortopassi, K.A., Hanser, S.F., McCowan, B., Ali, A.M. and Kirschel, A.N.G. (2011). Acoustic monitoring in terrestrial environments using microphone arrays: applications, technological considerations and prospectus. Journal of Applied Ecology, 48: 758-767. https://doi.org/10.1111/j.1365-2664.2011.01993.x.

Bristow, Thomas G., McHugh, N.M., Heward, C. J., Jenkins, D. L., Newson, S E., Snaddon, J. L. (2022) Vocal individuality measures reveal spatial and temporal variation in roding behaviour in Woodcock (Scolopax rusticola), Ibis, 10.1111/ibi.13176, **165**, 3, (959-973).

Chen, T., & Guestrin, C. (2016). XGBoost: A scalable tree boosting system. *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 785-794.

Christensen, T.K., Fox, A.D., Sunde, P. *et al.* (2017). Seasonal variation in the sex and age composition of the woodcock bag in Denmark. *Eur J Wildl Res* **63**, 52. https://doi.org/10.1007/s10344-017-1114-5

Dogan, H. (2025). Selected annotations of the Eurasian Woodcock (Scolopax rusticola) calls in Xeno-canto recordings from 2023 and 2024 [Data set][Version 1]. Zenodo. https://doi.org/10.5281/zenodo.15048227

Engler, J., Bokämper, M., Hannabach, S., et al. (2025). From dusk till dawn: Ecoacoustic monitoring reveals wind energy impacts on roding activity in the European Woodcock (Scolopax rusticola). *Authorea.* DOI: 10.22541/au.174117130.00877062/v1

Ghani, B., Denton, T., Kahl, S., & Klinck, H. (2023). Global birdsong embeddings enable superior transfer learning for bioacoustic classification. *Scientific Reports, 13*(1), 22876.

Goëau, H., Glotin, H., Vellinga, W. P., Planqué, R., & Joly, A. (2020). Overview of LifeCLEF 2020 Bird Recognition Challenge: Few-shot and zero-shot species classification from audio recordings. *CEUR Workshop Proceedings*, 2696.

Grill, T., & Schlüter, J. (2017). Two convolutional neural networks for bird detection in audio signals. *Proceedings of the Detection and Classification of Acoustic Scenes and Events (DCASE) Workshop*.

Heward, C. J., Conway, G. J., Hoodless, A. N., Norfolk, D., Aebischer, Nicholas J., (2024). Population and distribution change of Eurasian Woodcocks Scolopax rusticola breeding in the UK: results of the 2023 Breeding Woodcock Survey, Bird Study, 10.1080/00063657.2024.2345272, (1-15).

Ibrahim AK, Zhuang H, Schärer-Umpierre M, Woodward C, Erdol N and Chérubin LM (2024). Fish Acoustic Detection Algorithm Research: a deep learning app for Caribbean grouper calls detection and call types classification. *Front. Mar. Sci.* 11:1378159.

Kahl, S., Wood, C. M., Eibl, M., & Klinck, H. (2021). BirdNET: A deep learning solution for avian diversity monitoring. *Ecological Informatics, 61*, 101236.

Kath, H., Serafini, P. P., Campos, I. B., Gouvêa, T. S., & Sonntag, D. (2024). Leveraging transfer learning and active learning for data annotation in passive acoustic monitoring of wildlife. *Ecological Informatics, 82*, 102710. https://doi.org/10.1016/j.ecoinf.2024.102710

Kershenbaum, A., Bowles, A. E., Fournet, M. E. H., Ford, J. K. B., & Garland, E. C. (2016). Acoustic sequences in non-human animals: A tutorial review and prospectus. *Biological Reviews, 91*(1), 13-52.

Kramer, H. A., Kelly, K. G., Whitmore, S. A., Berigan, W. J., Reid, D. S., Wood, C. M., Klinck, H., Kahl, S., Manley, P. N., Sawyer, S. C., & Peery, M. Z. (2024). Using bioacoustics to enhance the efficiency of spotted owl surveys and facilitate forest restoration. *Journal of Wildlife Management, 88*, e22533.

Lasseck, M. (2018). Audio-based Bird Species Identification with Deep Convolutional Neural Networks. In: CEUR Workshop Proceedings.

Lasseck, M. (2023). Bird Species Recognition using Convolutional Neural Networks with Attention on Frequency Bands. In: CEUR Workshop Proceedings.

Liaw, A., & Wiener, M. (2002). Classification and regression by RandomForest. *R News, 2*(3), 18-22.

Mac Aodha, O., Gibb, R., Barlow, K. E., Browning, E., Firman, M., Freeman, R., & Scott, B. (2018). Bat detective—Deep learning tools for bat acoustic signal detection. *PLOS Computational Biology, 14*(3), e1005995.

Michaud, F., Sueur, J., Le Cesne, M., Haupert, S. (2023). Unsupervised classification to improve the quality of a bird song recording dataset, *Ecological Informatics*, Volume 74.

Prieto, N., Tavecchia, G., Telletxea, I. *et al.* (2019). Survival probabilities of wintering Eurasian Woodcocks *Scolopax rusticola* in northern Spain reveal a direct link with hunting regimes. *J Ornithol* **160**, 329–336.

Robinson, D., Miron, M., Hagiwara, M., & Pietquin, O. (2024). NatureLM-audio: An audio-language foundation model for bioacoustics. *arXiv*. https://arxiv.org/abs/2411.07186.

Schally, G., Tóth, D., Márton, M., Bijl, H., Palatitz, P., Csányi, S., Maimela Modiba, M., Tharwat Mohamed Ibrahim, H., & Simon, B. (2024). The effect of soil parameters and earthworm abundance on the fine-scale nocturnal habitat use of the Eurasian woodcock (*Scolopax rusticola*). *Ecology and Evolution*, 14, e70136.

Sokolova, M., & Lapalme, G. (2009). A systematic analysis of performance measures for classification tasks. *Information Processing & Management*, 45(4), 427-437.

Sprengel E., Jaggi M, Kilcher Y, Hofmann T., (2016). Audio based bird species identification using deep learning techniques. In: *CEUR Workshop Proceedings*.

Sládeček, M., Pešková, L., Chajma, P., et al. (2023). Eurasian woodcock (Scolopax rusticola) in intensively managed Central European forests use large home ranges with diverse habitats, *Forest Ecology and Management*, Volume 550.

Stowell, D., Wood, M., Pamuła, H., Stylianou, Y., & Glotin, H. (2019). Automatic acoustic detection of birds through deep learning: The first Bird Audio Detection challenge. *Methods in Ecology and Evolution*, 10(3), 368-380.

Tolkova, I., Chu, B., Hedman, M., Kahl, S., & Klinck, H. (2021). Parsing birdsong with deep audio embeddings. *arXiv*. https://arxiv.org/abs/2108.09203.

Tuti, M., Rodrigues, T.M., Bongi, P., Murphy, K.J., Pennacchini, P., Mazzarone, V., Sargentini, C. (2023). Monitoring Eurasian Woodcock (Scolopax rusticola) with Pointing Dogs in Italy to Inform Evidence-Based Management of a Migratory Game Species, *Diversity*, doi:10.3390/d15050598, **15**, 5, (598).

Xie, J., Towsey, M., Roe, P., & Truskinger, A. (2020). Embedding learning for automated bird call classification. *IEEE Transactions on Multimedia, 22*(7), 1850-1862.

Young MA, Spahr TB, McEnaney K, Rhinehart T, Kahl S, Anich NM, Brady R, Yeany D and Mandelbaum R., (2024). Detection and identification of a cryptic red crossbill call type in northeastern North America. *Front. Bird Sci*. 3:1363995. doi: 10.3389/fbirs.2024.1363995

Williams, B., van Merriënboer, B., Dumoulin, V., Hamer, J., Triantafillou, E., Fleishman, A. B., McKown, M., Munger, J. E., Rice, A. N., Lillis, A., White, C. E., Hobbs, C. A. D., Razak, T. B., Jones, K. E., & Denton, T. (2024). Leveraging tropical reef, bird, and unrelated sounds for superior transfer learning in marine bioacoustics. *arXiv*. https://doi.org/10.48550/arXiv.2404.16436