Improved detection of bird vocalisations using BirdNET embeddings and machine learning

Hakan Dogan, *Research Scientist*

***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 algorithm that has been trained using a large dataset of community 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) calls is investigated. BirdNET embeddings are used as feature representations and classifiers are trained based on these features. A strongly labeled dataset is created manually by annotating 97 recent recordings (2023–2024) from Xeno-canto, extracting 501 positive segments and 2,505 negative segments. We also make use of a second dataset available from the literature. BirdNET was then evaluated on both of these datasets, achieving an average precision of 84.3% and 89.3%, respectively. To enhance the detection accuracy, three machine learning classifiers are trained, i.e. Support Vector Machine (SVM), Random Forest, and XGBoost. The results indicate a significant improvement in classification performance, with overall average precision scores reaching the values of 99.8%–100% for both cases, in comparison to the baseline performance. Hence, the present work demonstrates that a hybrid (two-stage) deep learning approach, where the embeddings from a bird audio model are leveraged with posterior classifiers and strongly labeled data, can be a very accurate method for the recognition of bird species.

***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 (Kahl et al. 2021), a deep neural network trained for bird sound classification, has gained prominence due to its robustness and extensive training on diverse avian vocalizations (Michaud et al. 2023, Funosas et al. 2024, Fairbairn et al. 2024).

The state of the art deep neural networks for bioacoustics utilize large amounts of data and generally classify the vocalisations of many species simultaneously (Sprengel et al. 2016, Lasseck 2018, Lasseck 2023, Kahl et al. 2021, Robinson et al. 2024). The problem is handled as learning representations from the data for multiple classes and the objective is to minimize the discrepancy between the target classes and the model predictions. 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 (McGinn et al. 2023, Kramer et al. 2024, Young et al. 2024, Haley et al. 2024). This indicates that, whilst such deep neural networks deliver very robust feature representations (embeddings) in audio data, their performance may be leveraged for an individual species or for a custom geographical location (Ghani et al. 2023, Tolkova et al. 2021, Lasseck 2024, Bayat & Işık 2020, Huus et al. 2025, Somervuo et al. 2023). 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 & Zhu 2023, 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 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. 2015, Heward et al. 2024, Bristow et al. 2022, Tuti et al. 2023, Christensen et al. 2017, Schaly et al. 2024, Prieto et al. 2019, Sládeček et al. 2023). This species perform short-duration display flights (roding) during twilight to attract mates and mark territories, and they rely heavily on their acoustic environment (Engler et al. 2025). During the roding, the male birds perform typically two to five low-frequency calls described as grunts (400–1500 Hz), which are followed by a ~0.2 seconds long high pitched squeak with a maximum frequency up to 20 kHz. These characteristic display flights are typically carried out over forest clearings and along forest edge lines, and the observation of them is the only practical method for systematic surveys, owing to the secretive nature of the species (Holderried et al. 2025).

In the current method, the BirdNET embeddings are used as feature vectors and a secondary classifier is trained. The objective of the second classifier is to increase the differentiation between the target and the non-target acoustic events; therefore, non-target sound samples are also included in the training dataset (Singer et al. 2024, Nolan et al. 2023). We utilize 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 two-stage deep learning methods for classification problems (Khan et al. 2019, Sandoval et al. 2019, Pawar et al. 2022, Rycyk et al. 2022). As application for the bird sound recognition, Xie and Zhu (2023) used convolutional neural network (CNN) embeddings with an SVM classifier to improve species detection in urban environments. Specifically for BirdNET, Márquez-Rodríguez et al. (2025) have coupled it with Random Forest models for the identification of multiple species in complex soundscapes. Lakdari et al. (2024) have applied the same methodology, for the identification of individual gibbons. Note that the representations from such models have also been utilized for cross-domain tasks 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. METHODS

A. Data preparation

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 (Joly et al. 2020, Ghani et al. 2024, Pérez-Granados 2023, Pérez-Granados 2023b).

Dataset 1:

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 is 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 that 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.

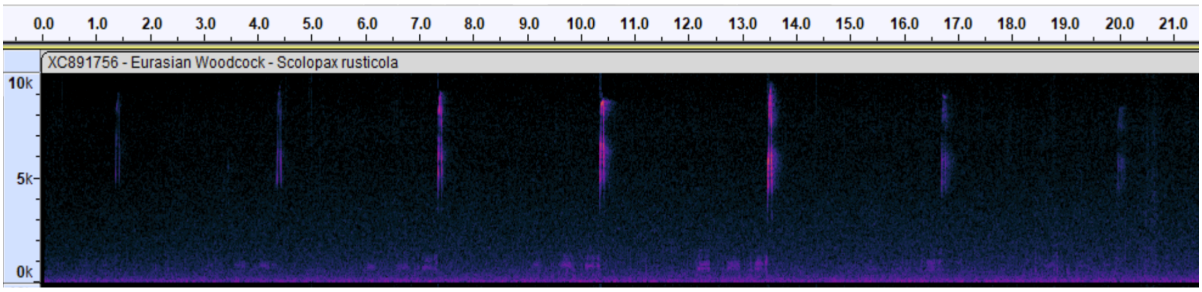

Fig 1. An example spectrogram for an Eurasian Woodcock recording (XC891756). During the flight of the bird, the grunt and the squeak calls as well as the changes to the amplitude strength are visible (created with Audacity).

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) and for ensuring the robustness of the evaluation metrics (Sokolova & Lapalme 2009).

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.

Dataset 2:

As another dataset, we have used the annotations from Holderried (2024), both for the assessment of BirdNet and for the algorithm development in the current article. The dataset includes 2514 annotations at audio-frame level. We have excluded a small portion of the data that are longer than 3-seconds in length, in order to keep the analysis at sample level, instead of file-based analysis. Note that the negative segments in Holderried (2024) have not been published publicly due to data privacy reasons. Alternatively, we concatenate the target samples from their dataset with the non-target recordings from the current study, which gives in total 2410 target samples and 2505 non-target samples, resulting in a balanced dataset in the current case.

B. 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. In Dataset 1, a total of 501 positive segments of Woodcock calls are contained, and to ensure robust model evaluation, a larger number of negative segments (n=2,505) is 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 probable scarcity of positive 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 a table that displays the cross-validation scores for each classifier across all 5 negative folds.

For Dataset 2, a balanced number of positive and negative samples is present. Therefore, it is not required to apply undersampling for the non-target samples. Nevertheless, a 5-fold cross-validation strategy is applied, i.e. 80% to 20% train-test split is created and the out-of-fold test score is computed for each split, similarly as Dataset 1.

III. RESULTS

A. BirdNET Predictions

To systematically evaluate BirdNET's performance for the detection of the Eurasian Woodcock calls, we conducted an inference test on both Dataset 1 and Dataset 2.

For Dataset 1, BirdNET achieves an average precision of 84.3%, which suggests that the model is quite 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. Also in Fig. 2, the recall vs. threshold curve is plotted, 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.

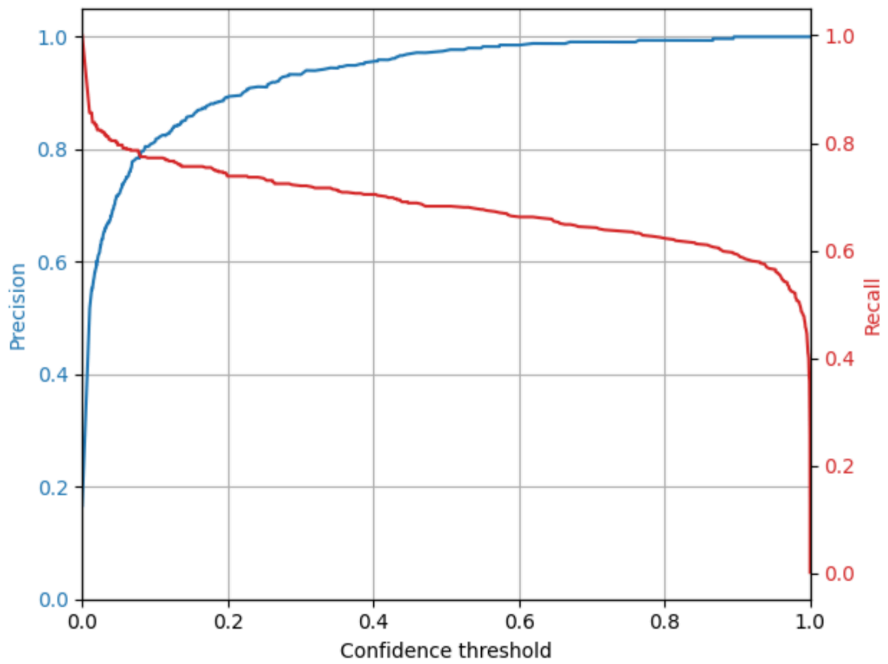

Fig 2. Precision vs. threshold and recall vs. threshold results for BirdNET on Dataset 1.

Figure 3 presents the precision vs. threshold and the recall vs. threshold results for Dataset 2. The BirdNET model exhibits quite high precisions for a wide range of confidence threshold values (blue line), whereas a recall value of 0.6 is achieved at the threshold value of 0.4. Overall, the algorithm gives an average precision value of 89.3 %. Holderried et al. (2025) have also reported similar values for the performance of BirdNET, i.e. 89.7 % true positive rate and 91.9 % accuracy for the occurrence of the species, and 81.7 % accuracy for the detection of roding events. This indicates that the negative samples in the current study are similar to theirs and that these datasets are both representative for environmental monitoring.

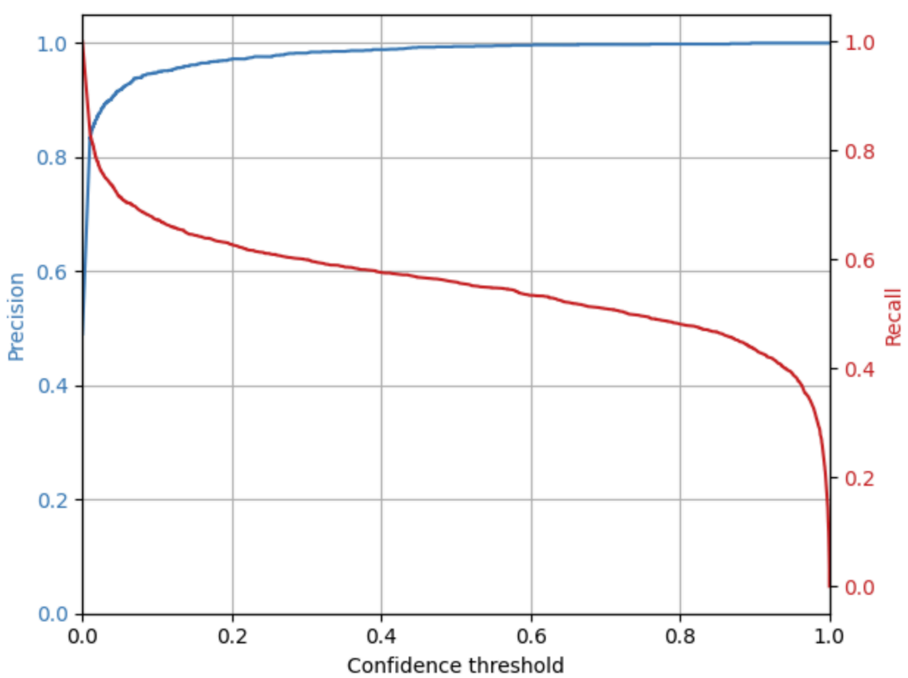

Fig 3. Precision vs. threshold and recall vs. threshold results for BirdNET on the Dataset 2.

Figure 4 presents the F1 scores related to the predictions of BirdNET for Dataset 1 and Dataset 2. The confidence (threshold) values of the BirdNET algorithm for specific species were investigated in several studies; see for example Tseng et al. 2025, Thompson et al. 2025, Pérez-Granados (2025). For the current work, a value of 0.815 is obtained for Dataset 1 as the maximum F1 score, and a maximum value of 0.833 is obtained for Dataset. The maximum F1 score occurs at the threshold value of 0.22 for Dataset 1 and at the threshold value of 0.01 for Dataset 2. The former is in accordance with the observations of Tseng et al. 2025, who report the optimal confidence values smaller than 0.35. However, the threshold value that maximizes the F1 score for Dataset 2 here, would yield a large number of detections and possibly a high amount of false positives in real monitoring scenarios (Engler et al. 2025).

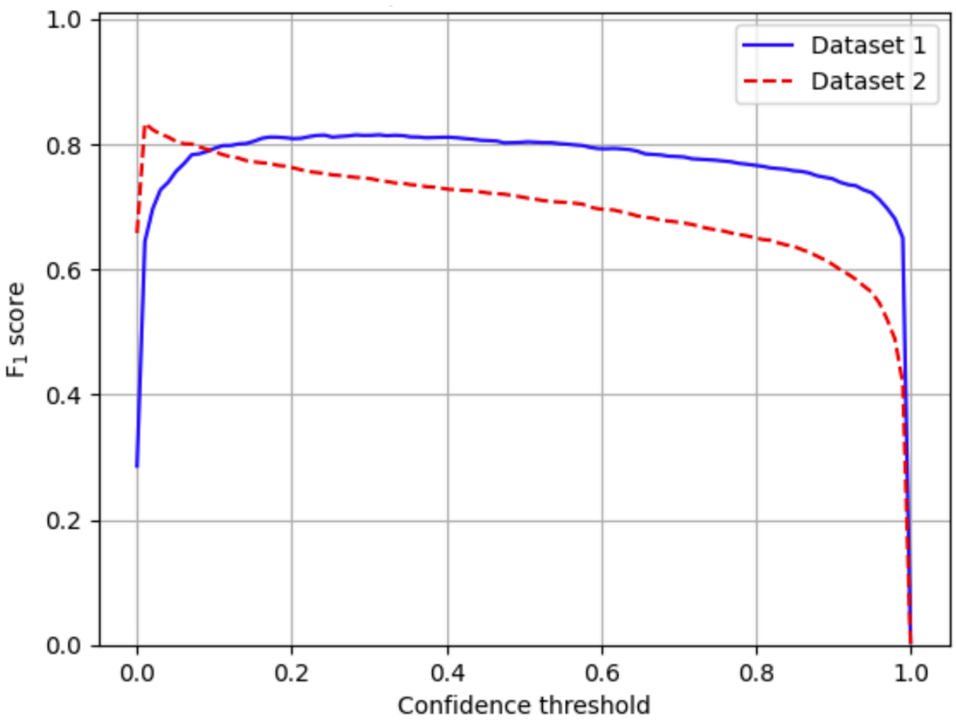

Fig 4. F1 results for the predictions of BirdNET on the Dataset1 (blue solid line) and Dataset 2 (red dashed line).

In addition to the above results, Engler et al. (2025) reports a precision value of 22.0 % for BirdNET's performance, i.e. 1475 true positives out of 6803 hits, where their analysis particularly focused on soundscapes near wind farm sites at dusk and dawn times. Based on these three studies (i.e. current work, Holderried et al. 2025 and Engler et al. 2025), BirdNET delivers overall strong performance characterized by high precision, 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.

B. Machine Learning Classifiers

In this section, the results of the posterior (second stage) classifiers are presented for both datasets - Dataset 1 and Dataset 2. As mentioned above, three classifiers are trained, i.e. SVM, Random Forest, and XGBoost, and a cross-validation methodology is applied in order to provide a robust analysis on the performance of the methods. Table 1 presents the mean accuracy values of the trained classifiers on Dataset 1, where the columns correspond to each of the negative folds and rows represent the employed machine learning methods. Table 2 shows the mean average precision values for the same simulations. Among the classifiers tested, SVM delivered the best performance, achieving an impressive mean accuracy value of 0.997 and average precision value of 1.0. 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 1. Mean Accuracy values of three different machine learning classifiers for Dataset 1

	Fold 0	Fold 1	Fold 2	Fold 3	Fold 4
SVM	0.996	0.997	0.995	0.997	0.998
Random Forest	0.980	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.

Table 2. Average precision values of three different machine learning classifiers for Dataset 1

	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

The mean accuracy values of the trained classifiers for Dataset 2 are presented in Table 3, where the columns correspond to each fold in the train-test split and rows represent the three machine learning methods - SVM, Random Forest and XGBoost. The results in Table 3 indicate an average mean accuracy value of 0.998 for SVM, 0.990 for Random Forest and 0.991 for XGBoost over all folds. Therefore, similar to the case in Dataset 1, SVM classifier delivers the highest accuracy. The corresponding average precision values on Dataset 2 for these algorithms are presented in Table 4. Averaging over all folds, we obtain a precision value of 1.0 for the SVM algorithm, 0.9996 for Random Forest and 0.9997 for the XGBoost. Therefore, all three algorithms exhibit a near-perfect precision value for a dataset with more than 5000 samples (Dataset 2), over all different train-test splits.

Table 3. Mean accuracy values of three different machine learning classifiers for Dataset 2

	Fold 0	Fold 1	Fold 2	Fold 3	Fold 4
SVM	0.99695	0.99797	0.99898	0.99695	0.99695
Random Forest	0.99084	0.99390	0.98983	0.98779	0.98678
XGBoost	0.99390	0.99084	0.98983	0.99288	0.98881

These results suggest that the two-stage deep learning methodology presented in the current paper performs very well on different datasets, train-test samples and almost in a classifier-agnostic way for the classification task introduced here. Overall, the feature space created by BirdNET embeddings could be leveraged very effectively by using posterior classification algorithms, providing a very accurate way of recognizing bird species in audio recordings.

Table 4. Average precision values of three different machine learning classifiers for Dataset 2

	Fold 0	Fold 1	Fold 2	Fold 3	Fold 4
SVM	1.00000	1.00000	1.00000	1.00000	1.00000
Random Forest	0.99966	0.99962	0.99970	0.99963	0.99946
XGBoost	0.99983	0.99988	0.99967	0.99952	0.99951

IV. CONCLUSION

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 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 achieved an average precision value of 84.3% for the Dataset 1 and 89.3 % for the Dataset 2 used in the current study, 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 there is currently a growing body of research data (Pérez-Granados et al. 2025, Morfi et al. 2019) in automated bioacoustic classification. Similarly, the strongly labeled dataset created during this work has also been made publicly accessible (Dogan, H. 2025).

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. DOI: 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. DOI: 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. DOI: 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 (Version 1)[Data set]. Zenodo. DOI: 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

Fairbairn, A. J., Burmeister, J.-S., Weisser, W. W., & Meyer, S. T. (2024). *BirdNET is as good as experts for acoustic bird monitoring in a European city*. bioRxiv. DOI: 10.1101/2024.09.17.613451

Funosas, D., Barbaro, L., Schillé, L., Elger, A., Castagneyrol, B., & Cauchoix, M. (2024). Assessing the potential of BirdNET to infer European bird communities from large-scale ecoacoustic data. *Ecological Indicators, 164*, 112146. doi: 10.1016/j.ecolind.2024.112146

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

Ghani, B., Kalkman, V. J., Planqué, B., Vellinga, W.-P., Gill, L., & Stowell, D. (2024). Generalization in birdsong classification: Impact of transfer learning methods and dataset characteristics. *arXiv* preprint. DOI: 10.48550/arXiv.2409.15383

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*.

Haley, S. M., Madhusudhana, S., & Branch, C. L. (2024). Comparing detection accuracy of mountain chickadee (*Poecile gambeli*) song by two deep-learning algorithms. *Frontiers in Bird Science,* 3:1425463. DOI: 10.3389/fbirs.2024.1425463

Heward, C. J., Hoodless, A. N., Conway, G. J., Aebischer, N. J., Gillings, S., & Fuller, R. J. (2015). Current status and recent trend of the Eurasian Woodcock *Scolopax rusticola* as a breeding bird in Britain. *Bird Study*, *62*(4), 535–551. DOI: 10.1080/00063657.2015.1092497

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. DOI: 10.1080/00063657.2024.2345272, (1-15).

Huus, J., Kelly, K. G., Bayne, E. M., & Knight, E. C. (2025). HawkEars: A regional, high-performance avian acoustic classifier. *Ecological Informatics, 87*, 103122. DOI: 10.1016/j.ecoinf.2025.103122

Holderried, P. (2024). Labelled acoustic dataset of roding Eurasian Woodcock (Scolopax rusticola). In Ibis (1.0.0). Forstliche Versuchs- und Forschungsanstalt Baden-Württemberg. DOI: 10.5281/zenodo.10930931

Holderried, P., Duschmalé, H., Günther, D., Isenberg, L. and Coppes, J. (2025), Essential steps for establishing a large-scale passive acoustic monitoring for an elusive forest bird species: the Eurasian Woodcock (*Scolopax rusticola*). Ibis, 167: 543-561. https://doi.org/10.1111/ibi.13373

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.

Joly, A., Goëau, H., Kahl, S., Deneu, B., et al. (2020). Overview of LifeCLEF 2020: A System-Oriented Evaluation of Automated Species Identification and Species Distribution Prediction. In: Arampatzis, A., *et al.* Experimental IR Meets Multilinguality, Multimodality, and Interaction. CLEF 2020. Lecture Notes in Computer Science(), vol 12260. Springer, Cham. https://doi.org/10.1007/978-3-030-58219-7_23.

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., Blumstein, D.T., Roch, M.A., Akçay, Ç., Backus, G., Bee, M.A., Bohn, K., Cao, Y., Carter, G., Cäsar, C., Coen, M., DeRuiter, S.L., Doyle, L., Edelman, S., Ferrer-i-Cancho, R., Freeberg, T.M., Garland, E.C., Gustison, M., Harley, H.E., Huetz, C., Hughes, M., Hyland Bruno, J., Ilany, A., Jin, D.Z., Johnson, M., Ju, C., Karnowski, J., Lohr, B., Manser, M.B., McCowan, B., Mercado, E., III, Narins, P.M., Piel, A., Rice, M., Salmi, R., Sasahara, K., Sayigh, L., Shiu, Y., Taylor, C., Vallejo, E.E., Waller, S. and Zamora-Gutierrez, V. (2016). Acoustic sequences in non-human animals: a tutorial review and prospectus. Biol Rev, 91: 13-52. doi: 10.1111/brv.12160

Khan FA, Gumaei A, Derhab A, Hussain A. (2019). A novel two‑stage deep learning model for efficient network intrusion detection. *IEEE Access*, 7:30373‑30385. https://doi.org/10.1109/ACCESS.2019.2899721.

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.

Lakdari, M. W., Ahmad, A. H., Sethi, S., Bohn, G. A., & Clink, D. J. (2024). Mel-frequency cepstral coefficients outperform embeddings from pre-trained convolutional neural networks under noisy conditions for discrimination tasks of individual gibbons. *Ecological Informatics, 80*, 102457. https://doi.org/10.1016/j.ecoinf.2023.102457

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.

Lasseck, M. (2024). Improving Bird Recognition using Pseudo-Labeled Recordings from the Target Location. In: CEUR Workshop Proceedings.

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

Márquez-Rodríguez, A., Mohedano-Munoz, M., Marín-Jiménez, M. J., et al. (2025). A Bird Song Detector for improving bird identification through Deep Learning: a case study from Doñana. *arXiv*. doi: 10.48550/arXiv.2503.15576

McGinn, K., Kahl, S., Peery, M. Z., Klinck, H., & Wood, C. M. (2023). Feature embeddings from the BirdNET algorithm provide insights into avian ecology. *Ecological Informatics, 74*, 101995. doi:10.1016/j.ecoinf.2023.101995

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, 101952.

Morfi V, Bas Y, Pamuła H, Glotin H, Stowell D. (2019). NIPS4Bplus: a richly annotated birdsong audio dataset. PeerJ Computer Science 5:e223. DOI: 10.7717/peerj-cs.223

Nolan, V., Scott, C., Yeiser, J.M., Wilhite, N., Howell, P.E., Ingram, D. and Martin, J.A. (2023), The development of a convolutional neural network for the automatic detection of Northern Bobwhite *Colinus virginianus* covey calls. Remote Sens Ecol Conserv, 9: 46-61. https://doi.org/10.1002/rse2.294

Pawar, Swati P., Talbar, Sanjay N. (2022). Two-Stage Hybrid Approach of Deep Learning Networks for Interstitial Lung Disease Classification, *BioMed Research International*, 7340902, 1-10. https://doi.org/10.1155/2022/7340902

Pérez-Granados, C. (2023), BirdNET: applications, performance, pitfalls and future opportunities. Ibis, 165: 1068-1075. DOI: 10.1111/ibi.13193

Pérez-Granados, C. (2023b), A First Assessment of Birdnet Performance at Varying Distances: A Playback Experiment. *Ardeola* 70(2), 257-269. DOI: 10.13157/arla.70.2.2023.sc1

Pérez-Granados, C., Etxebarría, J. M., Darras, K. F. A. et al. WABAD: A World Annotated Bird Acoustic Dataset for Passive Acoustic Monitoring, 03 January 2025, Preprint (Version 1), Research Square. DOI: 10.21203/rs.3.rs-5729784/v1

Pérez-Granados, C. (2025), Birdnet Confidence Scores Decrease with Bird Distance from the Recorder: Revisiting Pérez-Granados (2023). *Ardeola* 72(2), 149-159. DOI: 10.13157/arla.72.2.2025.fo1

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.

Rycyk, A., Bolaji, D. A., Factheu, C., & Takoukam, A. K. (2022). Using transfer learning with a convolutional neural network to detect African manatee (*Trichechus senegalensis*) vocalizations. *JASA Express Letters, 2*(12), 121201. DOI: 10.1121/10.0016543

Robinson, D., Miron, M., Hagiwara, M., & Pietquin, O. (2024). NatureLM-audio: An audio-language foundation model for bioacoustics. *arXiv*. doi: 10.48550/arXiv.2411.07186

Sandoval, C., Pirogova, E., & Lech, M. (2019). Two-stage deep learning approach to the classification of fine-art paintings. *IEEE Access, 7*, 41770–41781. https://doi.org/10.1109/access.2019.2907986

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.

Singer, D., Hagge, J., Kamp, J., Hondong, H. and Schuldt, A. (2024), Aggregated time-series features boost species-specific differentiation of true and false positives in passive acoustic monitoring of bird assemblages. *Remote Sens Ecol Conserv*, 10: 517-530. https://doi.org/10.1002/rse2.385

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

Somervuo, P., Lauha, P., & Lokki, T. (2023). Effects of landscape and distance in automatic audio based bird species identification. *The Journal of the Acoustical Society of America, 154*(1), 245–254. DOI: 10.1121/10.0020153

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. DOI: 10.1016/j.foreco.2023.121489

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.

Thompson, M.C., Ducey, M.J., Gunn, J.S. and Rowe, R.J. (2025), A post-processing framework for assessing BirdNET identification accuracy and community composition. Ibis, 167: 530-542. DOI: 10.1111/ibi.13357

Tolkova, I., Chu, B., Hedman, M., Kahl, S., & Klinck, H. (2021). Parsing birdsong with deep audio embeddings. *arXiv*. DOI: 10.48550/arXiv.2108.09203

Tseng, S., Hodder, D.P. & Otter, K.A. Setting BirdNET confidence thresholds: species-specific vs. universal approaches. *J Ornithol* (2025). DOI: 10.1007/s10336-025-02260-w

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., & Zhu, M. (2023). Acoustic Classification of Bird Species Using an Early Fusion of Deep Features. *Birds*, *4*(1), 138-147. https://doi.org/10.3390/birds4010011

Young M. A. , Spahr T. B., McEnaney K., Rhinehart T., Kahl S, Anich N. M., 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