Article

Not peer-reviewed version

Bioacoustic Detection of Wolves Using AI (BirdNET, Cry-Wolf and BioLingual)

Johanne Holm Jacobsen * , Pietro Orlando , Line Østergaard Jensen , Sussie Pagh , Cino Pertoldi

Posted Date: 16 December 2025

doi: 10.20944/preprints202512.1432.v1

Keywords: acoustic monitoring; bioacoustics; *Canis lupus*; wolf howls

 doi:10.20944/preprints202512.1432.v1

Article

Bioacoustic Detection of Wolves Using AI (BirdNET, Cry-Wolf and BioLingual)

Johanne Jacobsen [1,†], Pietro Orlando [2,†], Line Østergaard Jensen [1,†], Sussie Pagh [1] and Cino Pertoldi [1,3,*]

[1] Department of Chemistry and Bioscience, Aalborg University, 9220 Aalborg, Denmark
[2] Department of Agriculture, Mediterranean University of Reggio Calabria, 89124 Reggio Calabria, Italy
[3] Aalborg Zoo, 9000 Aalborg, Denmark
* Correspondence: cp@bio.aau.dk; Tel.: +4599403604
† These authors contributed equally to this work.

Simple Summary

Assessment of wolf populations today relies on multiple time-consuming and resource-intensive methods, including DNA testing of feces and wolf kills and wolf observations on wildlife cameras. This study aimed to explore wolf howls as an alternative monitoring tool for wolves and to compare several AI methods against a baseline of manual registration for detecting and classifying wolf howls from audio recordings. The results show that AI-based methods like BirdNET, BioLingual, and Cry-Wolf achieved high detection rates (78.5%, 61.5%, and 59.6% recall, respectively), though they also produced a substantial number of false positives. Crucially, combining these AI methods yielded an impressive 96.2% recall for actual howls. The use of automated AI methods significantly reduced the time spent on analysis of recordings, enabling the processing of larger datasets with fewer resources. This study demonstrates how the integration of AI-driven acoustic analysis can act as a non-invasive and efficient method, holding the possibility of becoming a standard for monitoring wolf populations and many other animal species.

Abstract

Rising numbers of wolf populations make traditional, resource-intensive methods of wolf monitoring increasingly challenging and often insufficient. This study explores how wolf howls can be used as a new monitoring tool fro wolves by applying AI methods to detect and classify wolf howls from acoustic recordings, thereby improving the effectiveness of wolf population monitoring. Three AI approaches are evaluated: BirdNET, Yellowstone's Cry-Wolf project system, and BioLingual. Data were collected using SM4 audio recorders in a known wolf territory in Klelund Dyrehave, Denmark, and manually validated to establish a ground truth of 260 wolf howls. Results demonstrate that while AI solutions currently do not achieve the complete precision or overall accuracy of expert manual analysis, they offer tremendous efficiency gains, significantly reducing processing time. BirdNET achieved the highest recall at 78.5% (204/260 howls detected), though with a low precision of 0.007 (resulting in 28,773 false positives). BioLingual detected 61.5% of howls (160/260) with 0.005 precision (30,163 false positives), and Cry-Wolf detected 59.6% of howls (155/260) with 0.005 precision (30,099 false positives). Crucially, a combined approach utilizing all three models achieved a 96.2% recall (250/260 howls detected). This suggests that while AI solutions primarily function as powerful human-aided data reduction tools rather than fully autonomous detectors, they represent a valuable, scalable, and non-invasive complement to traditional methods in wolf research and conservation, making large-scale monitoring more feasible.

Keywords: acoustic monitoring; bioacoustics; *Canis lupus*; wolf howls

 doi:10.20944/preprints202512.1432.v1

1. Introduction

The wolf population in Denmark has grown from the first confirmed wolf in 2012 to approximately 50 individuals by 2025, representing a steady recolonization across multiple territories[1,2]. This recovery has created both conservation opportunities for ecosystem restoration and practical management challenges for Danish wildlife authorities [3,4].

Denmark's wolf monitoring has historically relied on a collaborative framework between the Danish Environmental Protection Agency (Agency for Green Land Conversion and Aquatic Environment) and Aarhus University. The traditional approach combines DNA analysis from scat samples and saliva collected from wounds of animal killed by wolves, and wildlife camera trapping, and citizen-reported sightings[5,6]. However, these intensive monitoring protocols were discontinued in 2025 due to escalating costs as wolf numbers increased, with DNA analysis now limited to approximately 100 samples annually [7].

The current monitoring strategy has shifted toward pack-based population estimates using a conversion factor of 7 individuals per confirmed pack, introducing uncertainty due to natural variation in pack sizes and the challenge of accurately identifying all active packs across Denmark's expanding wolf territories [7]. Given these limitations in traditional monitoring approaches, alternative methods are increasingly necessary. Acoustic monitoring presents a promising complement for wolf population assessment, offering advantages in being non-invasive, relatively inexpensive, and capable of covering large geographical areas efficiently[6,8].

Wolf howls can transmit over distances of up to 6-11 kilometers under optimal conditions [9] and contain valuable information about pack composition [10], territorial boundaries [11], and even individual identity [12]. Previous research in Denmark has demonstrated the feasibility of individual wolf identification through acoustic analysis. Larsen et al. (2022) [13]successfully used multivariate analysis to distinguish between individual wolves and even different wolf subspecies from howl recordings, achieving high classification accuracy for solo howls. This groundwork established that wolf vocalizations contain sufficient individual-specific characteristics for identification purposes, suggesting that acoustic monitoring could potentially contribute to individual-level population assessment—a critical component for accurate wolf management in Denmark.

However, the manual analysis of acoustic recordings remains laborious [14], creating a bottleneck in data processing that limits broader implementation. This efficiency bottleneck has prompted interest in automated analysis solutions.

Acoustic monitoring has been successful for detection of other wildlife species, particularly in birds and bats [15,16]. These established applications demonstrate the maturity and reliability of AI-driven bioacoustics analysis, providing a strong foundation for adaptation to wolf monitoring applications.

Recent advances in artificial intelligence (AI) technologies offer potential solutions to the acoustic analysis challenge [17]. AI-based methods for automated detection and classification of animal vocalizations have shown promising results across various species [18], though their application to wolf monitoring, particularly in European contexts, remains an actively developing area [14,19]. The integration of these technologies could significantly enhance monitoring efficiency while maintaining necessary accuracy for conservation management decisions [20].

This study aims to compare three AI-based methods (BirdNET, Cry-Wolf, and BioLingual) against traditional manual analysis to explore their effectiveness in wolf monitoring scenarios within Danish territories. Detection accuracy, processing efficiency, resource requirements, and potential applications in conservation management are compared and evaluated. By comparing established approaches with AI-driven acoustic monitoring, the study seeks to explore how integrating AI methods can serve as a cost-effective complement to Denmark's evolving wolf monitoring strategy.

 doi:10.20944/preprints202512.1432.v1

2. Materials and Methods

2.1. Introduction to Study Species

The target species for this research is the Eurasian wolf (*Canis lupus lupus),* a subspecies representing the Central European population to which Denmark's resident wolves belong. Since the return of wolves to Denmark in 2012, several territories have been established, with the species showing gradual population recovery across the country[7].

2.2. Study Area

Data collections took place between August 2021 and February 2022 at Nature reserve in Southern Jutland, Denmark. The nature reserve is a publicly accessible wildlife park situated in a mixed habitat consisting of plantation forests, natural woodland, and heathland areas. This territory has been continuously occupied by wolves since August 2020, starting with a breeding pair comprising the Danish-born male and German-born female. The pair has demonstrated successful reproductive behavior, with documented litters including four pups born in May 2021 and six pups in 2022[7]. Acoustic recordings were conducted using passive recording equipment deployed throughout the territory to capture natural vocal behaviors. In August 2020 the pair gave birth to four pups in 2021 and six pups in 2022. From August 21, 2021, to February 20, 2022, two autonomous recorders operated for a total of 826.5 hours.

2.3. Acoustic Data Collection

Song Meter SM4 acoustic recorders (Wildlife Acoustics Inc., Maynard, MA, USA) were used for recordings. The acoustic recorders have a sampling rate of 44.1 kHz and an amplitude resolution of 16 bits. One recorder was positioned on a tree and one on a fence within the territory, approximately 1.5–2 m above ground. The recorders were set to auto-record continuously from dusk till dawn (17:00–07:30) as this corresponds to the period when wolves are most vocally active, and recordings were saved to SD cards with 128 GB storage capacity. All audio files were saved in WAV format to preserve recording quality for subsequent acoustic analysis.

2.4. Datasets for Evaluation

For the comparison of all three AI methods, four datasets were used. The datasets were separated chronologically to correspond with the recording capacity of the SD cards and battery of the recording equipment. This division served two purposes: it accommodated the logistical constraints of data retrieval and storage, and it allowed for the comparison of model performance across distinct seasonal periods. These datasets were meticulously manually annotated for wolf howls, serving as the "ground truth" for the analysis.

Across 826.5 hours of audio, we confirmed 260 wolf howls:

- dataset 1 (26 October–7 November 2021), 188.5 hours and 102 howls
- dataset 2 (22 November–5 December 2021), 203 hours and 50 howls
- dataset 3 (7–20 February 2022), 203 hours and 8 howls
- dataset 4 (28 September–13 October 2021), 232 hours and 100 howls

2.5. Manual Annotation and Ground Truth

All acoustic data designated for evaluation underwent a rigorous process of manual annotation to establish the definitive 'ground truth' for wolf howl occurrences. During this process, each identified wolf's howl was precisely marked with its start and end times. Furthermore, to provide detailed contextual information, specific labels were assigned to each howl. These annotations captured relevant environmental or acoustic conditions present during the howl, such as rain, red deer vocalizations, or unclear/faint howls. This manual annotation served as the baseline against which the performance of all automated detection methods was compared.

 doi:10.20944/preprints202512.1432.v1

2.6. Automated Detection Methods

Figure 1 summarizes the full pipeline, tracing the process from acoustic data collection and manual annotation through model application and alignment to quantitative performance evaluation.

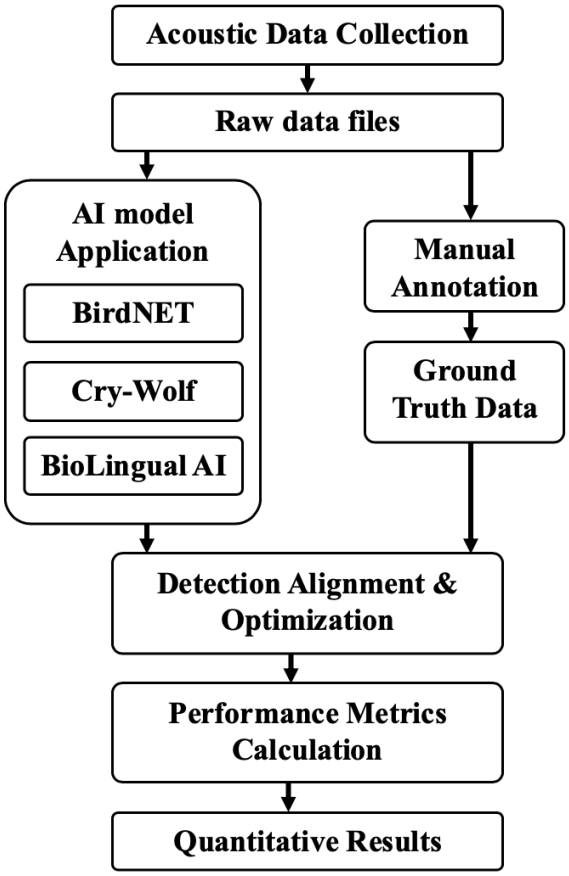

Figure 1. Illustrating the overall workflow of AI detection and evaluation.

Overview of AI Systems Tested

BirdNET was originally developed as a deep learning-based system for real-time identification of avian vocalizations, utilizing convolutional neural networks trained on extensive bird song databases. While primarily designed for bird species detection, BirdNET's extensive classification system includes over 6,000 classes, among which are several mammalian species, including the *grey wolf*. The model processes audio in fixed 3-second segments.

Cry-Wolf was developed specifically for wolf vocalization detection, with initial applications in Yellowstone National Park. The system operates through Kaleidoscope Pro software using targeted acoustic parameters: a frequency range of 200-750 Hz optimized for wolf howl fundamental frequencies, a duration range of 1.5-60 seconds, and FFT window size of 2.33 ms. This configuration allows the system to focus on acoustic characteristics most diagnostic of wolf howls while filtering out non-target sounds. Following acoustic feature extraction, the system employs traditional clustering methods to classify the detected segments. This approach represents a more conventional signal processing methodology compared to the deep learning-based models, relying on manually specified acoustic features rather than learned representations.

BioLingual represents a fundamentally different approach compared to the other models in this study, operating as a transformer-based language-audio model trained on the large-scale AnimalSpeak dataset containing over a million audio-caption pairs. Unlike traditional classification models that are trained on fixed class sets, BioLingual functions as a zero-shot classifier that can

 doi:10.20944/preprints202512.1432.v1

identify species calls without being explicitly trained on those specific classes. The model analyzes 10-second audio segments by finding the most similar label representation to an audio representation through cosine similarity calculations between audio embeddings and text embeddings of task label prompts. For this study, the model was provided with 350 classes derived from BirdNET's species list for Denmark, from which it selected the most appropriate label for each audio clip. The BirdNET classes were used to ensure consistency across methods.

2.7. Parameter Optimization and Alignment

Buffer Window Analysis for Detection Alignment

To account for temporal discrepancies between manual annotations and AI detections, buffer window analysis was conducted, testing 0, 3, and 10-second windows. The 10-second buffer was adopted as it significantly improved true positive capture (reducing missed detections by 23% across the three models and four datasets, while the associated increase in false positives for each model was below five < 5%, see in Appendix A Table A1).

Confidence Threshold Optimization (BirdNET)

BirdNET's confidence threshold was optimized through iterative testing across values from 0.001 to 0.5. The optimal threshold of 0.004 was selected based on optimal recall (78.5%; 204/260 howls), while minimizing false positives (28.7%).

Cry-Wolf Configuration

Cry-Wolf was run using its default pipeline with the Kaleidoscope Pro detector (200–750 Hz band-pass, 1.5–60 s event duration, 2.33 ms FFT window). To ensure comparability with the other models, we applied a uniform evaluation protocol: detections were aligned to the manual annotations using a symmetric 10-second tolerance window, and performance was scored with consistent definitions of true positives, false positives, and false negatives.

BioLingual Configuration and Confidence Threshold Optimization

BioLingual's confidence threshold was optimized through iterative testing across values from 0.9 to 0.95. The optimal threshold of 0.94 was selected based on optimal recall (61.5%; 160/260 howls) while minimizing false positives (30.1%)

2.5. Performance Metrics

Detection performance was quantified using precision, recall, and F1-score, computed from counts of true positives, false positives, and false negatives (Table 1).

Table 1. Glossary of Performance Metrics.

Metric	Definition
Precision	Proportion of positive detections that are actual wolf howls (TP / (TP + FP))
Recall	Proportion of actual wolf howls correctly detected by AI method (TP / (TP + FN)).
F1-score	Harmonic mean of precision and recall (2 * (Precision * Recall) / (Precision + Recall)).

3. Results

3.1. Overall Performance on Datasets

The results demonstrate significant variability across the three AI models within the fourdatasets. BirdNET achieved the highest recall (78.5%), successfully detecting the greatest proportion of actual wolf howls (Table 2).

 doi:10.20944/preprints202512.1432.v1

Table 2. Comprehensive Performance Overview for BirdNET, Cry-Wolf, and BioLingual on Datasets.

Dataset	Metric	BirdNET	Cry-Wolf	Biolingual
1	Precision	6.6% (76/1154)	2.5% (67/2678)	0.8% (36/4409)
	Recall	74.5% (76/102)	65.7% (67/102)	35.3% (36/102)
	F1-Score	0.121	0.048	0.016
2	Precision	3.2% (33/1022)	0.7% (30/4569)	0.5% (31/5808)
	Recall	66% (33/50)	60% (30/50)	62% (31/50)
	F1-Score	0.062	0.013	0.011
3	Precision	0.4% (6/1341)	0 (0/3278)	0.2% (4/2113)
	Recall	75% (6/8)	0% (0/8)	50% (6/8)
	F1-Score	0.009	0	0.004
4	Precision	0.3% (89/25460)	0.3% (58/19798)	0.5% (89/17924)
	Recall	89% (89/100)	58% (58/100)	89% (89/100)
	F1-Score	0.007	0.006	0.01

Table 3 reports model-level performance aggregated across the four datasets using a 10-s alignment window, presenting precision, recall, and F1-score; values in parentheses denote true positives over the corresponding denominator.

Table 3. Combined Performance Overview for BirdNET, Cry-Wolf, and BioLingual across Datasets.

Metric	BirdNET	Cry-Wolf	BioLingual
Precision	0.007 (204/28977)	0.005 (160/30254)	0.005 (155/30323)
Recall	78.5.61% (204/260)	59.6% (155/260)	61.5% (160/260)
F1-Score	0.014	0.01	0.01

3.2. Detailed Performance by AI System and Context

Impact of Environmental and Acoustic Conditions

Analysis of AI performance relative to manually annotated howl characteristics reveals system-specific strengths and vulnerabilities. An analysis of the labels of each howl revealed how environmental and acoustic conditions affected detection accuracy for the methods (Table 4).

 doi:10.20944/preprints202512.1432.v1

Table 4. AI Performance by Howl labels.

Note Type	Occurrences	BirdNET (Detected/Total)	Cry-Wolf (Detected/Total)	BioLingual (Detected/Total)
No label	32	28/32 (87.5%)	18/32 (56.3%)	15/32 (46.9%)
Rain	8	7/8 (87.5%)	6/8 (75.0%)	2/8 (25.0%)
Red deer	9	8/9 (88.9%)	7/9 (77.8%)	7/9 (77.8%)
Unclear	5	4/5 (80.0%)	2/5 (40.0%)	3/5 (60.0%)
Total	**54**	**47/54 (87.0%)**	**33/54 (61.1%)**	**27/54 (50.0%)**

This analysis illuminates specific challenges faced by each AI system on the data, with BirdNET showing consistent high performance across all acoustic conditions (Table 5).

Table 5. Performance pooled across the four datasets, evaluated with a 10-s symmetric alignment window.

Model	Recall (%)	Precision	False Positives (%)	Strenghts
Birdnet	78.5	0.007	28.773	87.5% during rain; 88.9% with deer rutting calls; 80% on unclear howls
Crywof	59.6	0.005	30,168	75% during rain; 77.8% with red deer sounds
BioLingual	61.5	0.005	30,094	Matched BirdNET's 89% recall on Dataset 4; 77.8% with red deer; 60% on unclear howls; detected 50% where Cry-Wolf had 0%

Across all three models, the principal limitation is very low precision (about 0.5–0.7%), yielding roughly 141–195 false positives per true positive and imposing a substantial verification workload. Cry-Wolf further shows instability—zero recall in one dataset and weak detection of faint howls—indicating sensitivity to recording conditions and temporal variation. BioLingual performs poorly in rain (25% detection) and is inconsistent even for clear howls (46.9%), with mixed outcomes under interference. BirdNET, despite high recall, is chiefly constrained by precision. Taken together, these weaknesses hinder scalability in heterogeneous soundscapes and motivate stricter post-processing, more conservative thresholding, and context-aware filtering to suppress false positives without eroding Recall.

4. Discussion

This study compared three AI-based methods—BirdNET, BioLingual, and Cry-Wolf—for detecting wolf howls in acoustic recordings from Dyrehave, Denmark. The recall rates observed were 78.5% for BirdNET, 61.5% for BioLingual, and 59.6% for Cry-Wolf. When combining the methods, the recall rate notably increased to 96.2%. Although all three models generated substantial numbers of false positives, they demonstrated strong potential as tools for human-assisted data reduction, especially when used together. This highlights their role not as fully autonomous detectors but as effective pre-screening solutions that can significantly streamline the manual review process, thus enhancing the feasibility of large-scale, non-invasive wolf monitoring efforts.

 doi:10.20944/preprints202512.1432.v1

The following sections explore the performance characteristics and practical implications of each AI detection method in greater detail, illustrating their complementary strengths and limitations in real-world acoustic environments.

4.1. Performance of AI Detection Methods: Interpretation and Practical Implications

Overall, BirdNET demonstrated the highest recall (78.5% across 260 howls), identifying 204 out of 260 manually confirmed howls. This positions BirdNET as the most suitable choice for initial screening in projects where the priority is to identify most potential wolf howls. Its strength in recall is evident across various environmental conditions noted in the ground truth data. For instance, BirdNET achieved 87.5% detection for howls under rainy conditions and with red deer (88.9%) in the background. Even for howls labeled as 'very unclear' howls, BirdNET detected 50% (11/22). This strong recall, particularly in diverse noisy contexts, reinforces its utility for capturing a broad range of vocalizations.

In contrast Cry-Wolf generally exhibited lower overall recall, detecting 155 out of 260 howls (59.6% recall). Like BirdNET, Cry-Wolf also presented notably low precision across all datasets (0.005 overall), indicating a substantial generation of false positives. The model's performance varied significantly with acoustic conditions; for instance, Cry-Wolf entirely missed all 8 howls in Dataset 3 (0% recall for that subset) and detected only 40% of faint howls.

BioLingual exhibited a slightly higher overall recall than Cry-Wolf, detecting 160 out of 260 howls (61.5% recall). However, like Cry-Wolf and BirdNET, BioLingual also presented consistently low precision across all datasets (0.005 overall), leading to a high number of false positives requiring human review. Its performance was also sensitive to environmental interferences, particularly performing poorly on howls occurring during rain (25% detection) and with bird background sounds. Despite these limitations, BioLingual demonstrated unique strengths in specific categories. Notably, it matched BirdNET's high recall in Dataset 4 (89% for both models), and in the detection of very faint howls (50% for both models) and faint howls in windy conditions (83.3% for both models).

This indicates BioLingual's potential for distinguishing the characteristics of howls even in faint recordings.

When all three detection methods were combined – and a positive detection from at least one method was considered – the cumulative detection rate improved significantly. From the 260 manually annotated howls, 250 were detected by at least one of the AI models. The higher combined detection rate (96.2%) strongly underscores the power of a multi-model approach in maximizing true positive identification. The observed performance patterns of the individual models, particularly in their recall-precision trade-offs and varied sensitivity to background noise, proved predictable in relation to their underlying architecture and training biases. A key finding is that there are no inherent trade-offs or disadvantages when running multiple AI models on the same dataset; instead, this strategy allows researchers to leverage the complementary strengths of each tool, optimizing for different research priorities (e.g., maximizing detection vs. minimizing manual review).

4.2. Broader Implications for Wolf Monitoring

AI-driven acoustic monitoring offers a powerful complement to traditional wolf monitoring methods, including DNA analysis from scat, camera trapping, and public observations[4].

The growing wolf population in Denmark[1,2] has made traditional intensive monitoring methods costly and largely reduced by 2025 [5,6,8]. Consequently, the national strategy now estimates wolf numbers based on pack counts, using a fixed conversion factor (e.g., 7 individuals per pack). This approach, however, introduces significant uncertainty due to variable pack sizes and challenges in identifying all active packs. Given these limitations and budget constraints that restrict intensive methods like DNA analysis to approximately 100 samples annually[7], AI-driven acoustic monitoring offers a promising, non-invasive, and cost-effective complement for wolf population assessment [6,7]. By integrating acoustic data with other monitoring streams like GPS telemetry or scat collection, it can optimize resource allocation by prioritizing areas for more intensive methods,

 doi:10.20944/preprints202512.1432.v1

and, in time, may also enable the identification of individual wolves and puppies. AI-driven acoustic monitoring can significantly support this evolving national strategy across several key areas:

Enhanced Presence/Absence Detection & Territoriality: AI-driven acoustic surveys can provide continuous monitoring over large areas. This is especially useful for detecting wolf presence in new or suspected territories, thereby triggering targeted follow-up with DNA collection or camera traps. This aligns with the need for efficient data collection given reduced DNA sampling budgets. It can also confirm continued occupancy in known territories with less reliance on frequent genetic re-identification and contribute data to map activity hotspots within territories.

Supporting Pack-Based Estimation: While AI howl detection, even with strong recall, doesn't directly identify individuals or precisely count pups, it can help confirm the presence of multiple vocalizing animals, suggestive of pack activity, thus supporting the identification of areas likely to contain packs. The detection of vocalizations labeled 'pup' howls by all models in our dataset, while limited, suggests the potential for AI to detect pup vocalizations, identifiable by distinct acoustic characteristics such as higher frequency energy[9] and shorter signals than adults [21], which would confirm breeding activity, a key element in defining a "pack" for national estimates.

Efficiency and Resource Allocation: Automated detection drastically reduces the human effort required to examine vast amounts of audio data. While manual validation of AI-flagged events is still necessary given the overall low precision (F1-scores ranging from 0.01 to 0.014), the ability to focus review only on flagged clips is far more efficient than manual listening to all recordings.

Non-Invasive Data Collection: Acoustic monitoring is entirely non-invasive, avoiding any disturbance to the animals, which is a significant advantage for sensitive species. Additionally, wolves may be monitored without entering private grounds.

4.3. Future Development Opportunities

Current AI systems, while effective for howl detection, have limitations that present significant opportunities for future improvement and expanded application:

Improving Precision: A critical area for future AI development is the significant improvement of precision to reduce false positives. The very low precision values observed in this study (overall 0.007 for BirdNET, 0.005 for BioLingual and Cry-Wolf) highlight that despite good recall, the models generate a vast number of false positives. While the 'cost' (human screening time)of reviewing a single false positive is minimal, the sheer volume can still be substantial, limiting scalability, especially in acoustically diverse wild environments[19,22]. Reducing this false positive burden will make these tools even more efficient and widely adoptable for large-scale monitoring efforts [23].

Full Analysis of Additional Datasets: A crucial next step for this research involves conducting a full manual annotation and comprehensive comparative AI analysis on more datasets. This will confirm the preliminary observations from and allow for the generalization of our findings across varied Danish wolf habitats, providing a more robust understanding of model performance in different environmental contexts.

Individual Identification: While individual wolf identification was beyond the scope of this study, substantial research evidence indicates this represents a promising possibility for future AI development. The 'from the same wolf ground truth note, while only occurring once, was detected by BirdNET, BioLingual and BirdNET, indicating that distinguishing individual vocalizations is possible. Larsen et al. (2022) [13] established individual wolf identification through acoustic analysis, achieving high classification accuracy using lonely howls. Supporting research demonstrates that wolf packs maintain distinctive, stable vocal signatures over time [24,25], and individual wolves possess unique vocal characteristics suitable for identification [26]. Future AI systems could incorporate sophisticated "acoustic fingerprinting" techniques, analyzing complex acoustic parameters for individual recognition.

Howl Type Differentiation: The methodology can be further developed to distinguish different howl types, such as howls, growls, barks, whines, and whimpers [27]. Our detection of a 'pup' howl across methods suggests this capability. Young wolves vocalize less frequently and produce shorter

 doi:10.20944/preprints202512.1432.v1

signals than adults [21], and their howls concentrate acoustic energy at higher frequencies [10]. This differentiation is crucial for reproductive monitoring [10,28] and assessing reproductive success [21]. BioLingual, with its language model foundation, shows promise for this nuanced differentiation of howl types.

Environmental Context and Human Impact: The data included howls influenced by various environmental sounds such as rain, very noisy, very windy and airplanes. While BirdNET maintained high recall in many of these conditions, the performance of other models varied significantly, indicating the impact of environmental factors. AI models could be further developed to analyze whether animals reduce vocalization after human noise events, linking to human disturbance and nocturnality patterns[29]. Environmental factors, including landscape, topography, vegetation cover, and weather, significantly affect sound wave propagation, influencing detected frequencies and localization accuracy [9,14,25,26]. Future models could integrate these environmental variables to improve detection and interpretation, as wolves may modulate their vocalizations based on environmental conditions[26].

Optimal Monitoring Periods: Understanding the ecological and behavioral context of wolf howling is essential for efficient recording schedules. Wolf howls primarily serve territory defense, intra-pack contact, and social bonding purposes[10,13,25,30,31]. Howling activity exhibits clear seasonal and circadian patterns, with peak intensity during July through October, coinciding with pup rearing[5,13,14,27,32], and predominantly at night, especially from dusk to early morning[29,31]. These insights inform optimal recording schedules for maximizing detection efficiency and minimizing the number of false detections.

Applicability to Other Species: The underlying AI methods, particularly BirdNET, are versatile and applicable to numerous other animal species for efficient and accurate identification across vast audio datasets [22,23], demonstrating the broader utility of this technology beyond wolf monitoring.

With recall rates of 59.6% to 78.5% for individual AI systems, and a combined detection rate of over 96%, their cost-effectiveness and efficiency gains through reduced manual review time make them invaluable tools. This highlights that a cost-effective system, even without perfect accuracy, can provide substantial benefits for large-scale monitoring efforts.

5. Conclusions

This study evaluated the effectiveness of monitoring wolves by their howls using three AI-based methods—BirdNET, BioLingual, and Cry-Wolf—for detecting wolf howls in passive acoustic recordings from Klelund Dyrehave, Suthern Jutland, Denmark. BirdNET achieved the highest recall of 78.5%, 61.5% for BioLingual, and 59.6% for Cry-Wolf. The results demonstrate that while none of the models are currently suitable as standalone detectors due to their high false positive rates, their primary value lies as highly effective human-aided data reduction tools. A combined approach of the AI methods dramatically improved overall detection to 96.2% recallThis study demonstrates that detecting wolf howls can serve as an effective new tool for identifying wolf presence. By applying AI models to classify calls and reduce manual review time, large-scale, non-invasive monitoring of wolf populations becomes significantly more feasible.

Author Contributions: Conceptualization, J.J., P.O. and L.Ø.J.; methodology, J.J., P.O. and L.Ø.J; software, J.J., P.O and L.Ø.J; validation, J.J., P.O. and L.Ø.J; formal analysis, J.J., P.O. and L.Ø.J; investigation, J.J., P.O. and L.Ø.J; resources, J.J., P.O. and L.Ø.J; data curation, J.J., P.O., L.Ø.J, C.P. and S.P.; writing—original draft preparation, J.J., P.O., L.Ø.J, C.P and S.P.; writing—review and editing, C.P. and S.P.; visualization, J.J., P.O., L.Ø.J; supervision, C.P. and S.P.; project administration, C.P. and S.P.; funding acquisition, C.P. and S.P. All authors have read and agreed to the published version of the manuscript.

Funding: This study was supported by the Aalborg Zoo Conservation Foundation AZCF: Grant number 07-2025 and The APC was funded by AZCF: 07-2025.

 doi:10.20944/preprints202512.1432.v1

Institutional Review Board Statement: The following scientific study did not require an Institutional Review Board Statement, in accordance with current regulations. Not applicable.

Ethical Considerations: Ethical approval was not required for this study and is therefore not applicable. The work relied exclusively on passive acoustic monitoring and non-invasive behavioral observations. Not applicable.

Data Availability Statement: Data are available from the first author on request.

Acknowledgments: We thank Klelund Dyrehave for providing access to the study site. We also thank Hanne Lyngholm Larsen for her technical assistance and to Line Østergaard Jensen for allowing us to use her data.

Conflicts of Interest: The authors declare no conflicts of interest.

Abbreviations

The following abbreviations are used in this manuscript:

AI	Artificial Intelligence
AZCF	Aalborg Zoo Conservation Foundation
SM4	Song Meter SM4 (Bioacustic recorder)
EU	European Union
FFT	Fast Fourier Transform
CNN	Convolutional Neural Network
FN	False Negative
FP	False Positive
Ff	Fundamental frequency
PAM	Passive Acoustic Monitoring
TN	True Negative
TP	True Positive

Appendix A

Table A1. Ground truth howls with labels and percentage detected for BirdNET, BioLingual and CryWolf.

Note Type	Occurrences	BirdNet (TRUE / Total)	BioLingual (TRUE / Total)	CryWolf (TRUE / Total)
(No Note)	74	68 (91.9%)	49 (66.2%)	68 (91.9%)
Other sound + rain	2	2 (100%)	0 (0%)	2 (100%)
Other sound at the end	1	1 (100%)	0 (0%)	1 (100%)
Other sound two times	1	1 (100%)	0 (0%)	1 (100%)
Another animal at the end	1	1 (100%)	0 (0%)	0 (0%)
bird	2	0 (0%)	0 (0%)	0 (0%)
birds	5	5 (100%)	0 (0%)	5 (100%)
From the same wolf	1	1 (100%)	1 (100%)	0 (0%)
Bark from dog	1	1 (100%)	1 (100%)	1 (100%)
Hoarse	1	1 (100%)	1 (100%)	1 (100%)
Puppy	1	1 (100%)	1 (100%)	1 (100%)
Red deer at the end	1	1 (100%)	1 (100%)	0 (0%)
Red deer	5	5 (100%)	3 (60.0%)	5 (100%)
With bird	2	2 (100%)	0 (0%)	0 (0%)
With red deer	1	1 (100%)	0 (0%)	1 (100%)
With duck	2	2 (100%)	1 (50.0%)	2 (100%)

 doi:10.20944/preprints202512.1432.v1

Very noisy	3	1 (33.3%)	0 (0%)	0 (0%)
Very faint/unclear	22	11 (50.0%)	11 (50.0%)	6 (27.3%)
Very windy	1	1 (100%)	1 (100%)	0 (0%)
Maybe other sound at the same time	1	1 (100%)	1 (100%)	1 (100%)
Maybe other sounds at the same time	1	1 (100%)	1 (100%)	1 (100%)
Rain	20	20 (100%)	7 (35.0%)	18 (90.0%)
Rain + other sound at the end	1	1 (100%)	0 (0%)	1 (100%)
Rain + red deer	1	1 (100%)	1 (100%)	1 (100%)
Rain + red deer	1	1 (100%)	0 (0%)	1 (100%)
Strong/clear howl	10	10 (100%)	5 (50.0%)	10 (100%)
Two wolves	1	1 (100%)	1 (100%)	1 (100%)
Two wolves?	1	1 (100%)	1 (100%)	1 (100%)
Wolf?	1	1 (100%)	1 (100%)	0 (0%)
Faint/unclear	39	33 (84.6%)	24 (61.5%)	25 (64.1%)
Unclear + airplane	1	0 (0%)	0 (0%)	1 (100%)
Faint/unclear + bark from dog	1	1 (100%)	1 (100%)	1 (100%)
Faint/unclear + red deer	1	1 (100%)	1 (100%)	1 (100%)
Faint/unclear + red deer	6	5 (83.3%)	5 (83.3%)	4 (66.7%)
Red deer in the beginning	1	1 (100%)	1 (100%)	1 (100%)
Total Rows	**260**	**204 (78.5%)**	**160 (61.5%)**	**155 (59.6%)**

Appendix B

Table A2. Glossary.

Acoustic fingerprinting	A technique for identifying unique acoustic characteristics or patterns in vocalizations that can distinguish between individual animals, similar to how human fingerprints are unique identifiers.
Amplitude resolution	The precision with which the magnitude of sound waves is digitally represented, measured in bits (e.g., 16- bit resolution provides 65,536 possible amplitude levels).
Audio embeddings	Mathematical vector representations of audio segments that capture acoustic features in a format suitable for machine learning analysis and comparison.
Bioacoustics	The scientific study of sound production, transmission, and reception in animals, including the ecological and behavioral contexts of animal vocalizations.
Buffer window	A time interval (in seconds) added before and after detected events to account for temporal discrepancies between automated detection and manual annotation boundaries.
CNN (Convolutional Neural Network)	A type of deep learning architecture particularly effective for analyzing data with spatial or temporal patterns, commonly used in image and audio processing.
Confidence threshold	A numerical cutoff point used by AI models to determine whether a detection is positive; detections with confidence scores above this threshold are classified as positive identifications.
Cosine similarity	A mathematical measure of similarity between two vectors, calculated as the cosine of the angle between them, commonly used to compare audio embeddings in AI models.
Deep learning	A subset of machine learning using neural networks with multiple layers to learn complex patterns and representations from data.

 doi:10.20944/preprints202512.1432.v1

F1- score	The harmonic mean of precision and recall, providing a single metric that balances both measures of classifier performance (formula: 2 × (Precision × Recall) / (Precision + Recall)).
False negative (FN)	An actual wolf howl that was not detected by the AI system (missed detection).
False positive (FP)	A detection by the AI system that was not actually a wolf howl (incorrect detection).
FFT (Fast Fourier Transform)	A mathematical algorithm that converts time-domain audio signals into frequency-domain representations, enabling analysis of the spectral content of sounds.
Fundamental frequency	The lowest frequency component of a periodic sound wave, which largely determines the perceived pitch of the sound
GPS telemetry	A wildlife monitoring technique using satellite- based Global Positioning System technology to track animal movements and locations in real-time.
Ground truth	The definitive, manually verified dataset against which automated detection systems are evaluated; represents the actual correct answers for comparison.
Language-audio model	An AI system that can process and understand relationships between textual descriptions and audio content, enabling cross-modal analysis.
Multivariate analysis	Statistical techniques that analyze multiple variables simultaneously to identify patterns, relationships, or classifications within complex datasets.
Passive acoustic monitoring	A non-invasive wildlife survey method using automated recording devices to capture animal vocalizations without human presence or intervention.
Precision	The proportion of positive detections that are actually correct (formula: True Positives / (True Positives + False Positives)).
Recall	The proportion of actual positive cases that were correctly detected (formula: True Positives / (True Positives + False Negatives)); also known as sensitivity.
Sampling rate	The frequency at which analog audio signals are converted to digital format, measured in Hz or kHz (e.g., 44.1 kHz means 44,100 samples per second).
Sonogram	A visual representation of sound showing frequency content over time, with frequency on the vertical axis, time on the horizontal axis, and intensity represented by color or darkness.
Spectral content	The distribution of energy across different frequencies within a sound signal, revealing the acoustic characteristics of vocalizations.
Transformer-based model	A neural network architecture that uses attention mechanisms to process sequential data, originally developed for natural language processing but adapted for audio analysis.
True negative (TN)	A correctly identified absence of a wolf howl (correct rejection).
True positive (TP)	A correctly identified wolf howl (correct detection).
Zero shot classification	An AI capability that allows models to classify or identify categories they were not explicitly trained on, using learned representations to generalize to new classes.

References

1. Rumlig Adfærd Af GPS-Mærket Ulv i Skjernreviret. https://dce.au.dk/fileadmin/dce.au.dk/Udgivelser/Notater_2023/N2023_21.pdf (Accessed on 11 December 2025).
2. Di Bernardi C, Chapron G, Kaczensky P, Álvares F, Andrén H, Balys V, et al. Continuing recovery of wolves in Europe. PLOS Sustain Transform 2025, *4(2):* e0000158. https://doi.org/10.1371/journal.pstr.0000158
3. Linnell, J.D.C.; Cretois, B. The Revival of Wolves and Other Large Predators and Its Impact on Farmers and Their Livelihood in Rural Regions of Europe; European Parliament, Policy Department for Structural and Cohesion Policies: Brussels, Belgium, 2018.
4. Directorate-General for Environment (European Commission); N2K Group EEIG; Blanco, J.C.; Sundseth, K. The Situation of the Wolf (*Canis lupus*) in the European Union: An In-Depth Analysis; Publications Office of the European Union: Luxembourg, 2023; ISBN 978-92-68-10281-7. https://data.europa.eu/doi/10.2779/187513 (Accessed on 11 December 2025).

5. Ražen, N.; Kuralt, Ž.; Fležar, U.; Bartol, M.; Černe, R.; Kos, I.; Krofel, M.; Luštrik, R.; Majić Skrbinšek, A.; Potočnik, H. Citizen Science Contribution to National Wolf Population Monitoring: What Have We Learned? *Eur J Wildl Res* 2020, *66*, pp. 46.
6. Marques, T.A.; Thomas, L.; Martin, S.W.; Mellinger, D.K.; Ward, J.A.; Moretti, D.J.; Harris, D.; Tyack, P.L. Estimating Animal Population Density Using Passive Acoustics. *Biological Reviews* 2013, *88*, pp. 287–309.
7. Olsen, K. e Sunde, P. , (2025). *Antallet af ulve i Danmark: Oktober 2012-februar 2025* , 20 p., Fagligt notat fra DCE – Nationalt Center for Miljø og Energi Vol. 2025 N. 26 https://dce.au.dk/fileadmin/dce.au.dk/Udgivelser/Notater_2025/N2025_26.pdf (Accessed on 11 December 2025).
8. Teixeira D, Maron M, van Rensburg BJ. Bioacoustic monitoring of animal vocal behavior for conservation. *Conservation Science and Practice*. 2019; 1:e72. https://doi.org/10.1111/csp2.72 (Accessed on 11 December 2025).ù
9. Arik Kershenbaum, Jessica L. Owens, Sara Waller; Tracking cryptic animals using acoustic multilateration: A system for long-range wolf detection. *J. Acoust. Soc. Am.* 1 March 2019; *145 (3),* pp. 1619–1628. https://doi.org/10.1121/1.5092973 (Accessed on 11 December 2025).
10. Palacios V, López-Bao JV, Llaneza L, Fernández C, Font E. Decoding Group Vocalizations: The Acoustic Energy Distribution of Chorus Howls Is Useful to Determine Wolf Reproduction. PLOS ONE, 2016, *11(5)*: e0153858. https://doi.org/10.1371/journal.pone.0153858 (Accessed on 11 December 2025).
11. R. McIntyre, J. B. Theberge, M. T. Theberge, D. W. Smith, Behavioral and ecological implications of seasonal variation in the frequency of daytime howling by Yellowstone wolves, *Journal of Mammalogy*, 29 May 2017, *98* (3), pp. 827–834, https://doi.org/10.1093/jmammal/gyx034 (Accessed on 10 December 2025).
12. Linhart, P., Mahamoud-Issa, M., Stowell, D. et al. The potential for acoustic individual identification in mammals. *Mamm Biol,* 2022, *102*, pp. 667–683. https://doi.org/10.1007/s42991-021-00222-2 (Accessed on 10 December 2025).
13. Larsen, H.L.; Pertoldi, C.; Madsen, N.; Randi, E.; Stronen, A.V.; Root-Gutteridge, H.; Pagh, S. Bioacoustic Detection of Wolves: Identifying Subspecies and Individuals by Howls. *Animals,* 2022, *12*, pp. 631. https://doi.org/10.3390/ani12050631 (Accessed on 11 December 2025).
14. Stähli, O., Ost, T., & Studer, T. Development of an AI-based bioacoustic wolf monitoring system. *The International FLAIRS Conference Proceedings*, 2022, *35*. https://doi.org/10.32473/flairs.v35i.130552 (Accessed on 10 December 2025).
15. Pérez-Granados, C. BirdNET: applications, performance, pitfalls and future opportunities. *Ibis,* 2023, *165*, pp. 1068-1075. *https://doi.org/10.1111/ibi.13193* (Accessed on 11 December 2025).
16. van Merriënboer, B.; Hamer, J.; Dumoulin, V.; Triantafillou, E.; Denton, T. Birds, Bats and beyond: Evaluating Generalization in Bioacoustics Models. *Front. Bird Sci,* 2024, *3*. https://www.frontiersin.org/journals/bird-science/articles/10.3389/fbirs.2024.1369756/full (Accessed on 11 December 2025).
17. Cauzinille, J.; Favre, B.; Marxer, R.; Rey, A. Applying Machine Learning to Primate Bioacoustics: Review and Perspectives. *American J Primatol* 2024, *86*, e23666. https://doi.org/10.1002/ajp.23666 (Accessed on 11 December 2025).
18. Guerrero, M.J.; Bedoya, C.L.; López, J.D.; Daza, J.M.; Isaza, C. Acoustic Animal Identification Using Unsupervised Learning. *Methods in Ecology and Evolution* 2023, *14*, pp. 1500–1514.
19. Sossover, D.; Burrows, K.; Kahl, S.; Wood, C.M. Using the BirdNET Algorithm to Identify Wolves, Coyotes, and Potentially Their Interactions in a Large Audio Dataset. *Mamm Res* 2024, *69*, pp. 159–165. https://doi.org/10.1007/s13364-023-00725-y (Accessed on 11 December 2025).
20. Wood, C.M.; Klinck, H.; Gustafson, M.; Keane, J.J.; Sawyer, S.C.; Gutiérrez, R.J.; Peery, M.Z. Using the Ecological Significance of Animal Vocalizations to Improve Inference in Acoustic Monitoring Programs. *Conservation Biology*, 2021, *35*, pp. 336–345. https://doi.org/10.1111/cobi.13516 (Accessed on 11 December).
21. Marti-Domken, B.; Sanchez, V.P.; Monzón, A. Pack Members Shape the Acoustic Structure of a Wolf Chorus. *acta ethol* 2022, *25*, pp. 79–87. https://doi.org/10.1007/s10211-021-00388-5 (Accessed on 11 December 2025).

22. Wood, C.M.; Kahl, S.; Barnes, S.; Van Horne, R.; Brown, C. Passive Acoustic Surveys and the BirdNET Algorithm Reveal Detailed Spatiotemporal Variation in the Vocal Activity of Two Anurans. *Bioacoustics* 2023, *32*(5), pp. 532-543. https://doi.org/10.1080/09524622.2023.2211544 (Accessed on 11 December 2025).
23. Wood, C.M.; Kahl, S. Guidelines for Appropriate Use of BirdNET Scores and Other Detector Outputs. *J Ornithol* 2024, *165*, pp. 777–782. https://doi.org/10.1007/s10336-024-02144-5 (Accessed on 11 December 2025).
24. Zaccaroni, M.; Passilongo, D.; Buccianti, A.; Dessì-Fulgheri, F.; Facchini, C.; Gazzola, A.; Maggini, I.; Apollonio, M. Group Specific Vocal Signature in Free-Ranging Wolf Packs. *Ethology Ecology & Evolution,* 2012, *24*(4), pp. 322–331. https://doi.org/10.1080/03949370.2012.664569 (Accessed on 11 December 2025).
25. Russo, C.; Cecchi, F.; Zaccaroni, M.; Facchini, C.; Bongi, P. Acoustic Analysis of Wolf Howls Recorded in Apennine Areas with Different Vegetation Covers. *Ethology Ecology & Evolution,* 2020, *32*, pp. 433–444.
26. Papin, M.; Pichenot, J.; Guérold, F.; Germain, E. Acoustic Localization at Large Scales: A Promising Method for Grey Wolf Monitoring. *Front Zool* 2018, *15*, 11. https://doi.org/10.1186/s12983-018-0260-2 (Accessed on 11 December 2025).
27. Passilongo, D.; Marchetto, M.; Apollonio, M. Singing in a Wolf Chorus: Structure and Complexity of a Multicomponent Acoustic Behaviour. *Hystrix It. J. Mamm.* 2017, 28(2), pp. 180-185. https://doi.org/10.4404/hystrix-28.2-12019 (Accessed on 11 December 2025).
28. Palacios, V.; Font, E.; García, E.J.; Svensson, L.; Llaneza, L.; Frank, J.; López-Bao, J.V. Reliability of Human Estimates of the Presence of Pups and the Number of Wolves Vocalizing in Chorus Howls: Implications for Decision-Making Processes. *Eur J Wildl Res* 2017, *63*:59, pp. 1-8. DOI 10.1007/s10344-017-1115-4.
29. Sunde, P.; Kjeldgaard, S.A.; Mortensen, R.M.; Olsen, K. Human Avoidance, Selection for Darkness and Prey Activity Explain Wolf Diel Activity in a Highly Cultivated Landscape. *Wildlife Biology* 2024, e01251. https://doi.org/10.1002/wlb3.01251 (Accessed on 11 December 2025)
30. Theberge, J.B.; Theberge, M.T. Triggers and Consequences of Wolf (*Canis Lupus*) Howling in Yellowstone National Park and Connection to Communication Theory. *Can. J. Zool.* 2022, *100*, pp. 799–809. https://doi.org/10.1139/cjz-2022-0043 (Accessed on 11 December 2025).
31. Harrington, F.H.; Mech, L.D. Howling at Two Minnesota Wolf Pack Summer Homesites. *Can. J. Zool.* 1978, *56*, pp. 2024–2028. https://doi.org/10.1139/z78-272 (Accessed on 11 December 2025).
32. Papin, M.; Aznar, M.; Germain, E.; Guérold, F.; Pichenot, J. Using Acoustic Indices to Estimate Wolf Pack Size. *Ecological Indicators,* 2019, *103*, pp. 202–211. https://doi.org/10.1016/j.ecolind.2019.03.010 (Accessed on 11 December).