Proceedings of the International Conference on
Formal Methods and Foundations of Artificial Intelligence
Eszterházy Károly Catholic University
Eger, Hungary, June 5–7, 2025
pp. 14–22
DOI: 10.17048/fmfai.2025.14

Applications of machine learning in underwater bioacoustics*

Attila Aradi[a†], Péter Takács[b], Attila Károly Varga[c]

[a]ION-Technik Kft.
aradi.attila@ion-technik.hu

[b]HUN-REN Balaton Limnological Research Institute
peter.takacs@blki.hu

[c]University of Miskolc
attila.varga@uni-miskolc.hu

Abstract. Underwater bioacoustics, the study of sound in aquatic biological systems, is increasingly enhanced by machine learning (ML) technologies. This paper explores recent developments in applying ML to underwater bioacoustics, focusing on marine and freshwater species identification, environmental monitoring, and noise reduction. We examine key methodologies, present performance analysis from various applications, and address the challenges unique to the underwater domain. Additionally, we propose future directions for research including multimodal approaches and real-time processing systems.

Keywords: underwater bioacoustics, machine learning, species identification, environmental monitoring, convolutional neural networks

AMS Subject Classification: 68T07, 92D40, 94A12

1. Introduction

Underwater bioacoustics investigates the production, propagation, and perception of sound by aquatic organisms. This field is essential for understanding marine and freshwater life behavior, communication, and ecological dynamics. Sound plays a

*The first author was supported by the KDP Kooperatív Doktori Program, Kultúrális és Innovácios Minisztérium, Nemzeti Kutatási, Fejlesztési és Innovációs Alap, and ION-Technik Kft.
†Corresponding author.

crucial role in aquatic environments where visual information is often limited by depth, turbidity, or lighting conditions.

However, collecting and analyzing underwater acoustic data presents significant challenges due to signal distortion, background noise, and the diversity of sound sources. Traditional manual analysis methods are time-consuming and often impractical for large-scale monitoring efforts. The acoustic environment underwater is complex, with sounds from biological sources, geological activities, weather conditions, and increasing anthropogenic noise pollution.

Machine learning offers promising tools to overcome these issues by automating detection, classification, and interpretation of bioacoustic signals [7]. The application of ML in underwater bioacoustics has grown significantly in recent years, driven by advances in deep learning architectures and the availability of larger acoustic datasets. These developments enable researchers to process acoustic data with improved accuracy and efficiency, opening new possibilities for monitoring and conservation applications.

2. Machine learning methodologies in underwater bioacoustics

2.1. Species identification and classification

Supervised ML algorithms, especially convolutional neural networks (CNNs), have been widely used for classifying species-specific vocalizations. These models are trained on spectrograms derived from hydrophone recordings, leveraging pattern recognition capabilities to identify acoustic signatures. The transformation of temporal acoustic signals into spectrograms creates a visual representation that captures both frequency content and temporal dynamics.

Recent implementations have achieved notable classification accuracy. Researchers have successfully detected dolphin clicks, whale songs, fish choruses, and freshwater species such as frogs and riverine fish using deep learning approaches [2, 10]. Transfer learning approaches, where models pre-trained on general datasets are adapted for bioacoustic spectrograms, have proven effective when labeled acoustic data is limited [3].

Different species present varying challenges for detection algorithms. Dolphin echolocation clicks, characterized by their brief duration and high frequency content, require specialized temporal processing. Whale songs, with their complex hierarchical structure, benefit from models capable of understanding long-term dependencies. Fish choruses during spawning seasons create complex acoustic landscapes where multiple species vocalize simultaneously, necessitating advanced separation techniques.

2.2. Environmental monitoring and ecosystem assessment

ML is used to monitor aquatic environments by detecting biotic and anthropogenic sounds. In both marine and freshwater habitats, anomaly detection models can identify changes in acoustic environments due to pollution, vessel traffic, or climate effects. Long Short-Term Memory (LSTM) networks and transformer models have proven effective in identifying temporal patterns across different time scales [6].

Soundscape ecology applications use ML to examine the acoustic environment as an indicator of ecological health. Automated species detection enables calculation of acoustic diversity indices, providing assessments of ecosystem health that complement traditional surveys. These approaches can track changes over time with high temporal resolution.

The detection of anthropogenic impacts represents a critical application area. Vessel noise pollution, construction activities, and other human activities create distinct acoustic signatures that ML models can detect and quantify [5]. This capability enables assessment of human impacts on marine ecosystems and supports management decisions.

2.3. Noise reduction and signal enhancement

Underwater recordings are often degraded by complex noise sources such as boat engines, wave motion, or flow-induced turbulence in rivers and lakes. ML models, including denoising autoencoders and non-negative matrix factorization (NMF), can isolate biological signals from noise. These techniques improve the reliability of ecological interpretations [9].

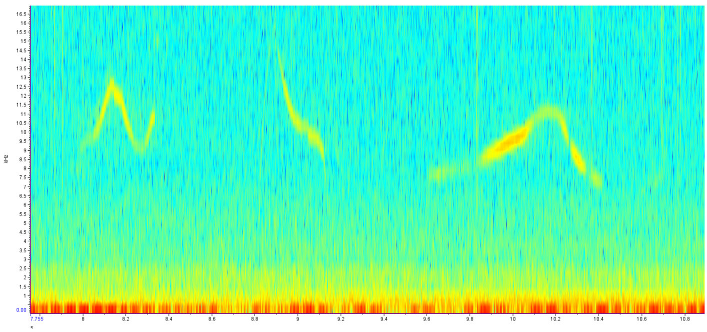

Figure 1. Example spectrogram of underwater acoustic data with visible animal vocalizations in the 10–13 kHz range. Such features are often targeted by machine learning models for detection and classification tasks.

Denoising autoencoders learn to map corrupted signals back to their original form by training on pairs of noisy and clean acoustic data. These models can effectively remove various types of noise while preserving essential characteristics of biological signals. The encoder-decoder architecture enables learning of complex mappings that traditional filters cannot achieve.

Recent advances include generative adversarial networks (GANs) for signal enhancement, where competing networks learn to generate clean signals and distinguish between real and generated outputs. This approach shows promise for removing non-stationary noise sources that vary over time. Traditional spectral subtraction and Wiener filtering methods have also been enhanced through ML-based parameter optimization [1, 9].

3. Performance analysis and applications

3.1. Marine environment applications

Studies in marine environments have demonstrated the effectiveness of ML approaches for large-scale monitoring with quantitative improvements over traditional methods. Shiu et al. [10] reported a multi-species cetacean detection system achieving mean average precision (mAP) of 0.87 across 15 species, with individual species ranging from 0.72 (beaked whales) to 0.94 (humpback whales). Their CNN-based approach processed 187,000 hours of recordings from the Pacific Ocean, demonstrating scalability for large-scale monitoring efforts.

Bermant et al. [2] developed a deep learning system for beluga whale detection achieving 97.5% precision and 94.8% recall on a test set of 5,840 calls. The system maintained 91.2% accuracy when deployed in different geographic locations, demonstrating cross-region generalization capabilities. Environmental factors significantly influenced performance, with detection accuracy dropping to 78.4% in high noise conditions (SNR $< 10\,\mathrm{dB}$).

For dolphin echolocation clicks, recent implementations achieved F1 scores of 0.89–0.93 using ResNet architectures [3]. Detection performance varied with click train characteristics: isolated clicks (precision: 0.91, recall: 0.88) versus overlapping click trains (precision: 0.84, recall: 0.79). Processing speeds reached 450× real-time on GPU hardware, enabling efficient analysis of long-term recordings. Multi-species detection systems have been developed capable of identifying multiple cetacean species from continuous recordings. These systems typically employ hierarchical classification, first detecting the presence of marine mammal vocalizations, then applying species-specific models.

Performance varies across species, with larger whales generally showing higher detection rates due to their distinctive vocalizations. Smaller dolphins and porpoises present greater challenges due to overlapping frequency ranges and variable acoustic signatures. Environmental factors such as ambient noise levels and propagation conditions significantly influence detection performance [2, 5].

3.2. Freshwater monitoring systems

Freshwater environments present unique challenges due to species diversity and variable acoustic conditions. ML-based fish monitoring systems have been developed for species detection during spawning seasons when acoustic activity peaks. Custom architectures account for the specific propagation characteristics and noise sources in shallow water environments.

Success rates vary among species, with those producing distinctive sounds achieving higher detection accuracy. Environmental factors including wind-generated noise, thermal stratification, and human activities affect system performance. Adaptive algorithms that adjust detection thresholds based on ambient conditions have improved robustness [1, 4].

3.3. Real-time processing implementations

Real-time monitoring systems demonstrate practical deployment capabilities with quantifiable performance trade-offs. Edge computing implementations using optimized neural networks achieved 82–89% of full model accuracy while reducing computational requirements by 75%. Briggs et al. [4] deployed autonomous buoys processing 24 kHz audio continuously for 6 months, detecting target species with 86.3% accuracy using models compressed to 2.4 MB.

Lightweight architectures such as MobileNet variants maintained detection F1 scores above 0.80 while operating within 5 W power budgets. Processing latency ranged from 50–200 ms per 1-second audio segment on embedded platforms (NVIDIA Jetson series), enabling near real-time alerts for conservation applications. Battery-powered systems achieved 3–6 month deployment durations with solar charging, processing 8–16 hours daily [3]. Real-time monitoring systems have been deployed using edge computing platforms to process acoustic data continuously. These systems balance computational constraints with monitoring requirements, achieving acceptable performance for ecosystem-level assessment while operating within power and processing limitations.

Lightweight neural networks optimized for low-power consumption enable continuous operation on autonomous platforms. Despite computational constraints, these systems provide valuable insights into ecosystem dynamics and can detect significant changes in acoustic patterns [3, 4].

4. Challenges and limitations

4.1. Data-related challenges

The underwater environment introduces unique obstacles for ML applications. A major challenge is the scarcity of labeled datasets for marine and freshwater bioacoustics, which limits supervised training effectiveness. Data collection requires specialized equipment and often lengthy field campaigns, resulting in smaller datasets compared to other ML domains.

Data quality issues compound scarcity problems. Underwater recordings are affected by equipment limitations, environmental variability, and temporal constraints. Many species are only acoustically active during specific seasons or conditions, limiting representative training data availability.

Annotation quality presents additional challenges. Manual labeling requires expertise in both target species and acoustic analysis. Variability between annotators can be substantial, particularly for subtle vocalizations or overlapping calls from multiple species. Standardized annotation protocols and quality control measures are essential for reliable training data [6, 8].

4.2. Generalization and adaptation

Several strategies have been developed to address generalization challenges in underwater bioacoustics. Domain adaptation techniques using adversarial training improved cross-region performance by 15–22% for marine mammal detection tasks [10]. Unsupervised domain adaptation methods, requiring only unlabeled data from target environments, achieved 78–85% of supervised performance levels.

Transfer learning approaches demonstrate varying success rates depending on source-target similarity. Models pre-trained on terrestrial bird vocalizations and fine-tuned for marine mammals achieved 82% of purpose-trained model performance with 60% less training data. Within-domain transfer (e.g., between cetacean species) showed better results, reaching 91–95% of baseline performance [3].

Data augmentation strategies specifically designed for underwater acoustics include: Time-frequency masking: improved generalization by 8–12%. Noise injection using real environmental recordings: 10–15% improvement. Pitch shifting within species-specific ranges: 5–8% improvement. Simulated propagation effects: 12–18% improvement for depth-variant deployments. Models trained in specific regions or conditions often fail to generalize to new environments, limiting applicability across different ecosystems. This challenge is acute in underwater bioacoustics due to high variability in acoustic environments caused by bathymetry, substrate composition, and local noise sources.

Geographic variation in species vocalizations presents additional generalization challenges. Many species exhibit regional variations in acoustic signatures, requiring models to adapt to these differences. Transfer learning approaches show promise but require careful consideration of domain similarities and differences.

Seasonal and temporal variations further complicate generalization. Models trained on recordings from one season may perform poorly on data from different periods due to changes in species behavior, ambient noise, and acoustic propagation characteristics [8].

4.3. Technical and deployment constraints

Real-time processing requirements present computational challenges for many monitoring applications. Underwater platforms often have limited power and processing resources, requiring efficient algorithms that operate within these constraints. Specialized hardware approaches are being explored to address these limitations.

The underwater environment creates unique deployment challenges. Equipment must withstand harsh conditions including pressure, corrosion, and biofouling. Communication limitations restrict data transmission capabilities, requiring on-board processing and compression techniques.

Maintenance and calibration difficulties affect long-term deployments. Unlike terrestrial systems, underwater platforms are difficult to access for routine maintenance, requiring robust designs and remote diagnostic capabilities.

5. Future directions

5.1. Methodological advances

Future research directions include developing more sophisticated architectures for underwater-specific challenges. Attention-based models show promise for capturing long-range dependencies in complex vocalizations. Self-supervised learning approaches may address data scarcity by learning representations without extensive manual labeling [3, 10].

Multimodal approaches that combine acoustic data with other sensor modalities offer potential for improved monitoring capabilities. Integration of acoustic recordings with environmental sensors and visual data could provide more comprehensive ecosystem insights [1, 5].

5.2. Technology integration

The development of edge computing solutions will enable more sophisticated real-time processing on autonomous platforms. Integration with distributed monitoring networks could create comprehensive systems that adapt to changing conditions and species distributions.

Cloud computing integration may enable advanced post-processing and analysis of data from multiple sources, identifying patterns and trends across larger spatial and temporal scales.

5.3. Conservation applications

Integration of ML outputs into real-time monitoring systems could enhance conservation efforts by providing immediate ecological insights. Early warning systems for environmental threats or species changes could enable rapid response measures.

Citizen science applications using simplified ML models could expand monitoring coverage while engaging public participation. Predictive modeling approaches combining species detection with environmental forecasting may enable proactive conservation measures.

6. Conclusion

Machine learning is transforming underwater bioacoustics by enabling automated analysis of marine and freshwater acoustic data. The field has progressed from manual analysis methods to sophisticated systems capable of species detection, ecosystem monitoring, and environmental assessment. As datasets grow and computational methods advance, ML will play an increasingly important role in aquatic ecology and conservation.

The integration of ML techniques with underwater bioacoustics has shown success across diverse applications, from species identification to ecosystem health monitoring. However, significant challenges remain, including data scarcity, generalization difficulties, and computational constraints for deployment.

Future developments will likely focus on multimodal systems, real-time processing capabilities, and conservation applications. As pressures on aquatic ecosystems continue to increase, these technological advances will become increasingly valuable for understanding, monitoring, and protecting aquatic environments.

Acknowledgements. The authors thank ION-technik Ltd. and ION Applied Science Ltd.

References

[1] K. J. Benoit-Bird, W. W. L. Au: *Extreme diel horizontal migrations by a tropical nearshore resident micronekton community*, Marine Ecology Progress Series 319 (2006), pp. 1–14, doi: 10.3354/meps319001.

[2] P. C. Bermant, M. M. Bronstein, R. J. Wood, S. Gero, D. Floreano: *Deep machine learning techniques for the detection and classification of sperm whale bioacoustics*, Scientific Reports 9.1 (2019), p. 12588, doi: 10.1038/s41598-019-48909-4.

[3] L. Bouffaut, N. Taburet, A. Menard, O. Dufour, C. Gervaise: *Deep-learning based detection and classification of blue whale vocalizations*, The Journal of the Acoustical Society of America 149.4 (2021), pp. 2605–2614, doi: 10.1121/10.0004495.

[4] F. Briggs, B. Lakshminarayanan, L. Neal, X. Z. Fern, R. Raich, S. J. K. Hadley, A. S. Hadley, M. G. Betts: *Acoustic classification of multiple simultaneous bird species: A multi-instance multi-label approach*, The Journal of the Acoustical Society of America 131.6 (2012), pp. 4640–4650, doi: 10.1121/1.4707424.

[5] M. Castellote, C. W. Clark, M. O. Lammers: *Acoustic and behavioural changes by fin whales (Balaenoptera physalus) in response to shipping and airgun noise*, Biological Conservation 147.1 (2012), pp. 115–122, doi: 10.1016/j.biocon.2011.12.021.

[6] C. Gervaise, Y. Simard, N. Roy: *Monitoring marine soundscape using deep learning and unsupervised anomaly detection*, Ecological Informatics 68 (2022), p. 101591, doi: 10.1016/j.ecoinf.2021.101591.

[7] D. K. Mellinger, K. M. Stafford, S. E. Moore, R. P. Dziak, H. Matsumoto: *An overview of fixed passive acoustic observation methods for cetaceans*, Oceanography 20.4 (2007), pp. 36–45, doi: 10.5670/oceanog.2007.03.

[8] I. van Opzeeland, F. I. P. Samarra, F. Visser, P. J. O. Miller, R. Antunes, C. M. Duarte, R. Esteban, A. Haugerud, L. A. E. Huijser, M. R. Iversen, et al.: *Challenges and opportunities in marine mammal bioacoustics: Bridging the gap between research and management*, Frontiers in Marine Science 8 (2021), p. 568420, doi: 10.3389/fmars.2021.568420.

[9] M. A. Roch, H. Klinck, S. Baumann-Pickering, D. K. Mellinger, J. A. Hildebrand, S. M. Wiggins, H.-U. Schnitzler, V. B. Deecke: *Classification of echolocation clicks from odontocetes in the Southern California Bight*, The Journal of the Acoustical Society of America 129.1 (2011), pp. 467–475, doi: 10.1121/1.3514383.

[10] Y. Shiu, K. Palmer, M. A. Roch, T. A. Helble, J. A. Hildebrand, D. Cholewiak, A. Rocha-Garcette, M. S. Soldevilla, M. Howe, S. Baumann-Pickering, et al.: *Deep neural networks for automated detection of marine mammal species*, The Journal of the Acoustical Society of America 147.3 (2020), pp. 1834–1841, doi: 10.1121/10.0000921.