INTERNATIONAL JOURNAL ON SMART SENSING AND INTELLIGENT SYSTEMS

Review | DOI: 10.2478/ijssis-2026-0077 Issue 1 | Vol. 19 (2026)

Sensor-Enabled Automated Bioacoustic Systems for Precision Livestock Management and Wildlife Welfare Monitoring: A Comprehensive Review

Sugandha Sharma*, Virender Kadyan and Shareef Babu Kalluri

School of Computer Science, UPES, Dehradun, Uttarakhand, India

*E-mail: sugandha.sharma@ddn.upes.ac.in

Received for publication March 02, 2026.

Abstract

Acoustic signals play a vital role in animal communication and offer a valuable, non-invasive means of assessing animal health and welfare. Insights from bioacoustic analysis allow livestock farmers to adopt sound-based monitoring systems that improve farm management while minimizing direct interaction with animals. The use of audio and acoustic sensors provides a practical and animal-friendly approach for monitoring large populations, enabling continuous, real-time observation that is less intrusive than conventional physiological methods. However, several challenges persist, including limited availability of high-quality datasets, variability in recording environments, and difficulties in interpreting model outputs, which collectively hinder large-scale deployment in farm settings. To address these limitations, the incorporation of advanced data augmentation strategies and multi-modal sensor integration has been suggested to enhance data diversity and improve classification performance. Current research efforts are focused on developing cost-effective, real-time, and robust bio-acoustic monitoring systems that are adaptable to varying farm conditions and species. Such systems have the potential to transform livestock management by enabling early detection of health and welfare issues, thereby supporting more sustainable and ethically responsible farming practices. This paper provides a comprehensive review of bioacoustic-based approaches for animal welfare monitoring, encompassing a wide range of techniques including individual identification, stress detection, estrus recognition, and respiratory condition analysis. Spanning over a decade of research, the current study underscores the evolving use of machine learning and deep neural networks, including convolutional and transformer-based models, in decoding complex animal vocalizations. This review develops a comprehensive taxonomy and provides a critical synthesis of sensor-enabled bioacoustic systems for livestock and wildlife monitoring. It further outlines major research gaps and practical deployment challenges, proposing a forward-looking framework to facilitate scalable and real-world implementation of these systems.

Keywords

sensor, sensor-based animal welfare monitoring, bioacoustics, animal welfare, acoustic monitoring, machine learning, deep learning, animal vocalization, non-invasive sensing, precision livestock farming

Open Access. Professor Subhas Chandra Mukhopadhyay. CC BY-NC-ND
© 2026 Sharma et al., This work is licensed under the Creative Commons Attribution-NonCommercial-NoDerivs 4.0 License https://creativecommons.org/licenses/by-nc-nd/4.0/

I. Introduction

Acoustics is the active production of sounds through vibrations from vocal folds, which manifests and expresses different inner states of an animal like physiological states like being in heat (estrus), hungry, tired, in pain, or stressed, emotional states such as fear, excitement, aggression, contentment, or distress and motivational or social states like warning others of danger, or call offspring [1]. The production of sounds like animal calls may occur impulsively or may be the outcome of independent external stimuli [2]. Acoustic signal plays a crucial role in delivering information between animals or humans [1]. Vocalization plays a very important means of communicating with other animals, along with intercommunication with the environment [2]. Animals have the ability to generate distinct patterns of vocal sounds in various conditions, such as fever, stress, hunger, estrus, and sickness [2]. The generated acoustic information helps the animals to understand each other or to convey information related to a particular situation like hunger, stress, an animal's individuality, age, gender, dominance state, estrus condition, or any medical emergency [3]. Like humans, animal sounds also vary due to various acoustic variability factors, such as gender, age, and environment, like forests, farms etc. [3].

A species' vocalization pattern transmits physiological, emotional, as well as individual-specific information, which further specifies that these vocalization patterns can be utilized as potential indicators to understand and learn about distinct animal conditions and welfare [4]. Sounds arising from non-communicative sources such as by-products of behavioral activities like resting or grazing, physical processes like breathing, or environmental factors such as weather, complement the information conveyed by bio-acoustic signals and can also reveal details about social behavior, reproductive status, and both internal and external interactions [5–9]. These sounds validate the value of analyzing animal vocal recordings. Recently, automated methods have been developed for investigating and classifying acoustic sounds, particularly in environmental biology, welfare, and conservation to automatically classify animal sounds for tasks such as species identification and population (animal-count) estimation [10]. Vocalization can be considered a non-intrusive approach to evaluate the welfare of animals when compared to other biophysical metrics. Despite this potential, these approaches haven't found extensive application in animal welfare monitoring. With targeted efforts directed toward computing advances and instrumentation, data collected from acoustic recording units (ARU) can be used to document the welfare state of subject animals [11]. ARU are automated, stand-alone devices designed to capture and store sounds from the environment over extended periods without the need for a human operator to be present.

Bio-acoustic monitoring may offer a distinctive possibility to surveil behavior at a more granular level [11]. Many studies in recent times have been reported to identify particular conditions like estrus, separation distress, respiratory disorders etc. in cattle using vocal-sound sensors to assure cattle's health and welfare without disturbing the designated farm-budget. These techniques are cost-effective, as they are inexpensive. Besides the cost, the sensors are more useful and animal-friendly since they gather data constantly without triggering any distress or uneasiness in the subject animals [12–17].

The present review offers the following major contributions:

1. It explicitly emphasizes the transition of bio-acoustic monitoring from experimental laboratory frameworks toward scalable, interpretable, and autonomous intelligent sensing ecosystems for next-generation livestock management and wildlife conservation.
2. By integrating technological, computational, and deployment-oriented perspectives within a unified review framework, the manuscript provides a broader and more critical understanding of current advancements, unresolved challenges, and future research opportunities in bioacoustic monitoring systems.

Section I outlines the background of the study and the methodology adopted for the literature review. Section II provides a comprehensive examination of bioacoustic research in animal welfare, covering applications such as stress detection, speaker identification, and estrus detection, along with an overview of various neural network architectures explored in this domain. Section III discusses the key challenges and practical considerations associated with real-world deployment of these systems. Finally, Section IV summarizes the study and highlights potential avenues for future research and development. This is further followed by References section.

Section a presents a background with the recent studies conducted in the area of bioacoustics, specifically in the case of bovines, aquatic species, avian species, swines, elephants, insects, and other species.

a. Background

This section highlights the bioacoustic studies conducted on various species of animals. The authors in [11] conducted on vocalization analysis of hens reflected that temperature changes caused stress in chickens. A study has been carried out in the direction of respiratory disease detection in broilers based on sneezing and coughing sounds [12, 13]. Liu et al. [14] performed an investigation that sought to detect avian influenza using chicken vocal cues.

In bovine and dairy cattle research, certain studies have harnessed audio data to detect estrus [15] as well as respiratory diseases [17, 18]. Research work has also been carried out in swines to detect abnormalities in them using their vocal signals. Early studies used the frequency and time domain features and showed the statistical significance in detecting the abnormalities in pig sounds [19–22]. These studies are aimed toward identifying disease-related coughs and stress-induced screams. A failure in detection of respiratory diseases in the early phase, stress-induced screams, and failing to identify aggressive behaviors among pigs that have been weaned will result in significant financial loss [23–26]. However, the latest investigations have incorporated machine learning (ML) approaches to enhance the accuracy of identifying specific conditions like coughing sounds and stress-induced screams in swines [23, 26]. Previous studies had certain limitations because they primarily focused on verifying whether animal vocalization data could be used to identify abnormal behaviors in environments such as barns and pigpens. Consequently, animal acoustics play an important role, as they provide valuable insights that can aid in detecting a wide range of irregularities in livestock. In the case of small cattle herds, accurate detection of estrus remains a major challenge that directly affects reproductive efficiency [27]. Although the relationship between bioacoustics and reproductive performance is still in its early stages of investigation, understanding this link through specific animal vocalizations is essential [2, 27–29]. Determining the estrus stage in dairy cattle at the right time plays a crucial role in enhancing milk production, livestock welfare, and remunerative profits to the dairy industry, including livestock farmers [3]. There is a dearth of automated monitoring systems that try to recognize the calls of individual cattle animals for specific requirements and respond accordingly [29–32]. In a small unorganized herd, where one livestock farmer breeds and raises five to six animals only, they are easily able to identify the calls of individual animals for particular needs even from a distant location and respond accordingly [29–32]. Sound has come to be regarded as a valuable, non-invasive source of information for monitoring and understanding animal health, welfare, and productivity. Livestock produce a wide range of vocalizations that reflect their physiological states and behavioral needs, offering a promising avenue for precision farming. The authors in [33] highlight that acoustic cues in livestock farming can serve as real-time indicators of stress, discomfort, or environmental challenges like an increase in the temperature of barn or leaving the stock for grazing in open farms, enabling farmers to improve animal welfare and management practices. Building on this premise, advances in artificial intelligence (AI) and ML have accelerated the emergence of automated solutions capable of interpreting these complex acoustic signals. For example, Kannan et al. [34] demonstrate how vocalization-based gender identification in chicks can be achieved through AI and ML techniques, underscoring the potential of data-driven approaches for early decision-making in poultry production. Beyond poultry, automated sound classification is also gaining importance in broader animal monitoring contexts. Cheema and Basker [35] describe how wildlife surveillance can benefit from robust animal sound classification, illustrating that acoustic analytics can extend from farm management to biodiversity conservation. In the domain of dairy farming, understanding reproductive physiology is crucial for enhancing breeding efficiency. Abebe et al. [36] emphasize that timely and accurate detection of estrus, combined with artificial insemination protocols, is a cornerstone in improving reproductive performance in cattle. Integrating acoustic monitoring with reproductive management strategies could thus provide a novel, non-invasive tool to complement traditional methods of estrus detection and herd fertility assessment. Taken together, these studies underscore a growing intersection between bioacoustics, animal reproduction, and intelligent computing. Harnessing animal vocalizations through artificial intelligence-driven (AI-driven) analysis offers the potential not only intended to elevate animal welfare and production but also to enable early, precise, and scalable interventions in both farming and wildlife conservation [33, 34].

Extensive research in bioacoustics has focused on speaker recognition, identification, and discrimination across a variety of animal groups, including large mammals such as elephants and bovines [37, 38]. In aquatic environments, classification techniques integrate temporal and spectral features to identify complex marine sounds, employing advanced hybrid

neural networks and deep learning (DL) models [39, 40]. These methods have greatly enhanced species detection amid noisy underwater conditions and large datasets, thereby facilitating ecosystem monitoring and biodiversity assessments [41, 42]. The use of transfer learning and data augmentation has further improved performance, particularly in small aquatic datasets, supporting more accurate classification of whale, fish, and dolphin vocalizations [43]. In this context, Li et al. [44] introduced the multi-scale temporal residual transformer (MT)-Resformer (multi-scale two-channel fusion network), a fusion network with multi-scale and two-channel design that effectively combines multi-resolution features and dual-channel inputs to enhance marine mammal call classification, demonstrating superior accuracy and robustness in challenging underwater acoustic environments.

Avian bio-acoustics leverages convolutional neural networks (CNNs) and ML models to classify bird species and individuals, frequently learning directly from raw audio waveforms. This has considerably improved multi-species classification accuracy, contributing meaningfully to ecological monitoring and conservation efforts [45–47]. Automated acoustic indices complement these techniques by enabling biodiversity assessments at the soundscape level, effectively distinguishing biological from anthropogenic sounds [48]. Expanding on these advances, Zianouka et al. [49] developed an AI–based system for identifying bird sounds specific to Belarus's bird species, demonstrating how region-specific datasets and ML models can enhance large-scale monitoring of local bird populations. Similarly, Nikshya et al. [50] applied neural network models developed for sound-based monitoring of the Himalayan monal and other rare avian species, highlighting the potential of advanced neural architectures to detect rare or cryptic species and thereby strengthen conservation strategies in challenging field environments.

In bovine studies, bio-acoustic research integrates feature extraction with DL architectures like CNNs to identify vocalizations and behavioral patterns, with on-animal acoustic sensors allowing real-time welfare monitoring and farm management [51–53]. Acoustic cues have also been employed to differentiate individual cattle, including Karan Fries cows [54]. Context-specific vocalization studies have explored the relationship between emotional states and sound, such as humming calls in Holstein heifers linked to weaning stress; this suggests that bio-acoustics can be a tool for welfare assessment [55]. For early disease detection in calves, efforts utilize automated analysis of cough sounds to identify bovine respiratory diseases in calves, while poultry respiratory conditions have been assessed via sound analysis in commercial broiler farms [4, 42].

Swine bioacoustic classification primarily targets health and welfare, employing vocal feature extraction methods, such as mel-frequency cepstral coefficients (MFCCs), formants, and energy metrics enhanced by neural networks and fusion strategies. These approaches are effective in disease detection—including respiratory syndromes—and reproductive monitoring, thus advancing precision farming [56–59]. Wild boar vocalization study highlights emotional call types distinct from domestic pigs, emphasizing the complexities of extrapolating findings across species [4, 48]. The study by Sukkuea and Akkajit [60] demonstrates the effectiveness of segmentation-based DL techniques for behavior recognition in group-housed pigs.

The calls of wild animals and mammals, including African elephants, have been classified with hidden Markov models (HMMs) trained using Baum–Welch algorithm and tested via the Viterbi algorithm, achieving accurate automatic classification of calls [42]. (Makarov, N., Savchenko, A., Zemtsova, I., Novopoltsev, M., Poyarkov, A., Viricheva, A., … & Hernandez-Blanco, J. A. (2025). Automated detection of wolf howls using audio spectrogram transformers. Scientific Reports, 15(1), 26641) proposed a DL model that distinguished wolf howls from other animal sounds. The potential for ML and DL in the recognition and monitoring of elusive or endangered species has been evident [4]. Building on these foundations, Geldenhuys and Niesler [61] employed deep neural architectures to automate the classification, detection, and segmentation of elephant calls and sub-calls, demonstrating improved precision in handling the acoustic complexity of low-frequency (LF) rumbles. Similarly, Muir et al. [62] developed methods for detecting and distinguishing vocal calls made by coppery titi monkeys in captivity, illustrating the adaptability of bio-acoustic monitoring to primate communication studies. In the context of carnivore monitoring, Makarov and Ivanov [63] introduced an audio spectrogram transformer framework for the automated detection of wolf howls, enabling robust recognition in variable field conditions. Extending these innovations to endangered mammals, Xu and Liao [64] proposed a hybrid Local–Global Transformer parallel network combined with long short-term memory (LSTM) to accurately recognize the calls of newborn giant pandas, highlighting the value of advanced DL models in conservation-driven acoustic monitoring [35, 36, 54, 65–67].

In the field of insect bio-acoustics, Vasconcelos et al. [68] suggested an efficient approach for mosquito species detection and classification by optimizing two-dimensional audio feature extraction for lightweight ecoacoustic applications. Their method demonstrated that carefully selected spectral representations can significantly improve the accuracy of species-level classification while keeping computational demands low, making it suitable for large-scale environmental monitoring and mobile or resource-constrained field deployments. Lee et al. [69] demonstrated a highly effective mel-frequency cepstral coefficients-linear discriminant analysis (MFCC-LDA) approach for automatic recognition of animal vocalizations for species like frogs and crickets. The MFCC-LDA method proved robust for both insect and amphibian sound classification, establishing a foundation for further bio-acoustic studies. This validated the potential of straightforward audio features and statistical classifiers across diverse taxa [70].

Besides stress detection and disease detection in animals, reproductive acoustic cues in cattle and dairy animals have also garnered attention in recent times. Decision support systems employing ML have been developed for estrus detection in buffaloes and cows, with acoustic parameters such as call duration, pitch, formants, and harmonic-to-noise ratios serving as key indicators of estrus phases [66, 67, 70, 71]. Increases in vocalization rates during estrus correspond with visual signs, suggesting vocal features as reliable reproductive biomarkers in livestock [16, 28, 54, 55, 72]. Recent advances widely apply DL models, including MnasNet and ensembles of CNNs, such as GoogleNet, AlexNet, and ResNet, along with Vision Transformers—to enhance classification across species, encompassing birds, whales, bats, cattle, and pigs [37, 42, 43, 51]. Despite these technological gains, research into dairy animal vocal behavior is still limited, and the multifaceted nature of acoustic signals complicates the decoding of animal conditions [73–78]. Continuing advancements in ML, data augmentation, and acoustic analysis promise to deepen understanding and application of bioacoustics in animal welfare, ecology, and precision farming [37, 55, 72, 79–81]. Indriasari et al. [82] highlighted the importance of knowledge management systems in supporting smallholder dairy farming, providing a foundation for integrating intelligent monitoring and decision-support technologies. Yazykbayev et al. [83] highlight sustainable feed production through waste valorization, supporting the broader goals of efficient and technology-driven livestock management systems.

Dataset scarcity remains a significant challenge in bioacoustic systems, particularly in livestock monitoring applications. The size of datasets employed in bioacoustic applications varies considerably based on the particular use case. For instance, estrus detection studies typically rely on relatively small datasets ranging from approximately 100 to 1,000 samples. On the other hand, illness detection tasks, such as cough sound analysis, usually use datasets that are medium-sized, with between 500 and 5,000 samples. Wildlife monitoring apps, especially those that use long-term passive acoustic recordings, sometimes have substantially bigger datasets, with more than 10,000 recordings in many cases. This difference in the size of the datasets shows how hard it is to create generalized models, and how important it is to have consistent, large-scale annotated datasets in bioacoustic research.

This review covers both livestock and wildlife monitoring, with a primary focus on livestock systems while highlighting transferable methodologies for wildlife applications. Unlike prior reviews, this work provides a unified analysis of sensor hardware, acoustic signal processing, and AI-based modeling, with a specific focus on deployment challenges and real-world feasibility in agricultural environments. Section b presents the review methodology used in this study.

b. Review methodology

This review follows the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) framework to ensure a transparent, systematic, and reproducible study selection process. The review was conducted through four sequential phases: identification, screening, eligibility assessment, and inclusion. The identification process included a comprehensive literature search that was performed across multiple databases, including Web of Science, Scopus, IEEE Xplore, and Google Scholar, using predefined keywords and Boolean expressions related to bioacoustics, animal sound analysis, livestock monitoring, wildlife monitoring, and ML. This process resulted in the identification of 151 studies. This was followed by a screening process where duplicate records were removed, and the remaining studies were screened based on titles and abstracts to assess relevance to the scope of sensor-enabled bioacoustic systems. Subsequently, eligible full-text articles were then evaluated against predefined inclusion and exclusion criteria. Only high-quality peer-reviewed journal and conference papers published between 2015 and

2025, with experimental validation, were considered. Studies lacking empirical results or falling outside the domain were excluded. Following the eligibility assessment, a total of 115 studies were selected for qualitative and detailed analysis.

b.i Literature search strategy

A comprehensive literature search was performed across four widely recognized scientific databases, namely Web of Science, Scopus, IEEE Xplore, and Google Scholar. The search was restricted to publications published between 2015 and 2025, reflecting the rapid advancements in sensor technologies, bioacoustic monitoring, and AI during this period.

The search strategy employed predefined keywords combined using Boolean operators (AND/OR) to maximize the retrieval of relevant studies. Representative search strings included: ("bio-acoustic" OR "acoustic monitoring" OR "animal vocalization") AND ("precision livestock farming" OR "livestock monitoring" OR "animal welfare") AND ("machine learning" OR "deep learning" OR "artificial intelligence"), as well as ("wildlife monitoring" OR "biodiversity monitoring") AND ("acoustic sensor" OR "internet of things" (IoT) OR "edge computing"). To minimize the possibility of omitting relevant publications, the reference lists of selected review articles and highly cited research papers were also manually examined for additional eligible studies.

b.ii Study screening and selection

The initial database search identified 151 records. Following the removal of duplicate publications, 135 unique studies were retained for title and abstract screening. During this phase, publications that were unrelated to sensor-enabled bioacoustic systems, animal acoustic analysis, or intelligent monitoring technologies were excluded. The remaining 125 full-text articles were subsequently evaluated against predefined eligibility criteria. After a detailed assessment, 115 studies satisfied all selection requirements and were included in the final qualitative synthesis. The complete workflow of the study selection process is presented in Figure 1.

b.iii Inclusion and exclusion criteria

To ensure consistency and scientific rigor, explicit inclusion and exclusion criteria were established prior to the literature screening process. To ensure a consistent and rigorous study selection process, predefined inclusion and exclusion criteria were established prior to the literature screening. Publications were considered eligible if they were peer-reviewed journal articles or conference papers published between 2015 and 2025, with a primary focus on sensor-enabled bioacoustic applications in precision livestock management, wildlife monitoring, or animal welfare assessment. Eligible studies were also required to employ acoustic signal processing in combination with ML, DL, or other AI techniques, and to provide experimental validation, field implementation, or quantitative performance evaluation of the proposed methodology. Publications were excluded when their scope did not align with automated bioacoustic sensing or intelligent animal monitoring. In addition, review papers, editorials, book chapters, dissertations, patents, and other non-peer-reviewed publications were omitted to ensure the inclusion of original research only. Studies lacking sufficient methodological detail or experimental validation were also removed, as were duplicate records and publications written in languages other than English.

b.iv Quality assessment

To enhance the reliability and credibility of the review, all eligible studies underwent a qualitative assessment based on methodological rigor and reporting quality. The evaluation considered several aspects, including the clarity of research objectives, adequacy of experimental design, description of sensor configurations, dataset characteristics, validation methodology, reproducibility of the proposed approach, and the practical significance of the reported outcomes. Preference was given to studies providing comprehensive experimental validation and clearly documented methodologies, thereby ensuring that the synthesized findings were supported by robust scientific evidence.

b.v Study classification framework

Following the selection process, the included studies were organized into four thematic categories based on their primary application: (i) Stress Detection, (ii) Speaker Identification, (iii) Estrus Detection, and (iv) Neural Networks for bioacoustics. This classification provided the structural framework for the subsequent sections of the review and enabled a systematic comparison of technological developments, performance trends, and research gaps across application domains. The selected publications were systematically classified to facilitate comparative

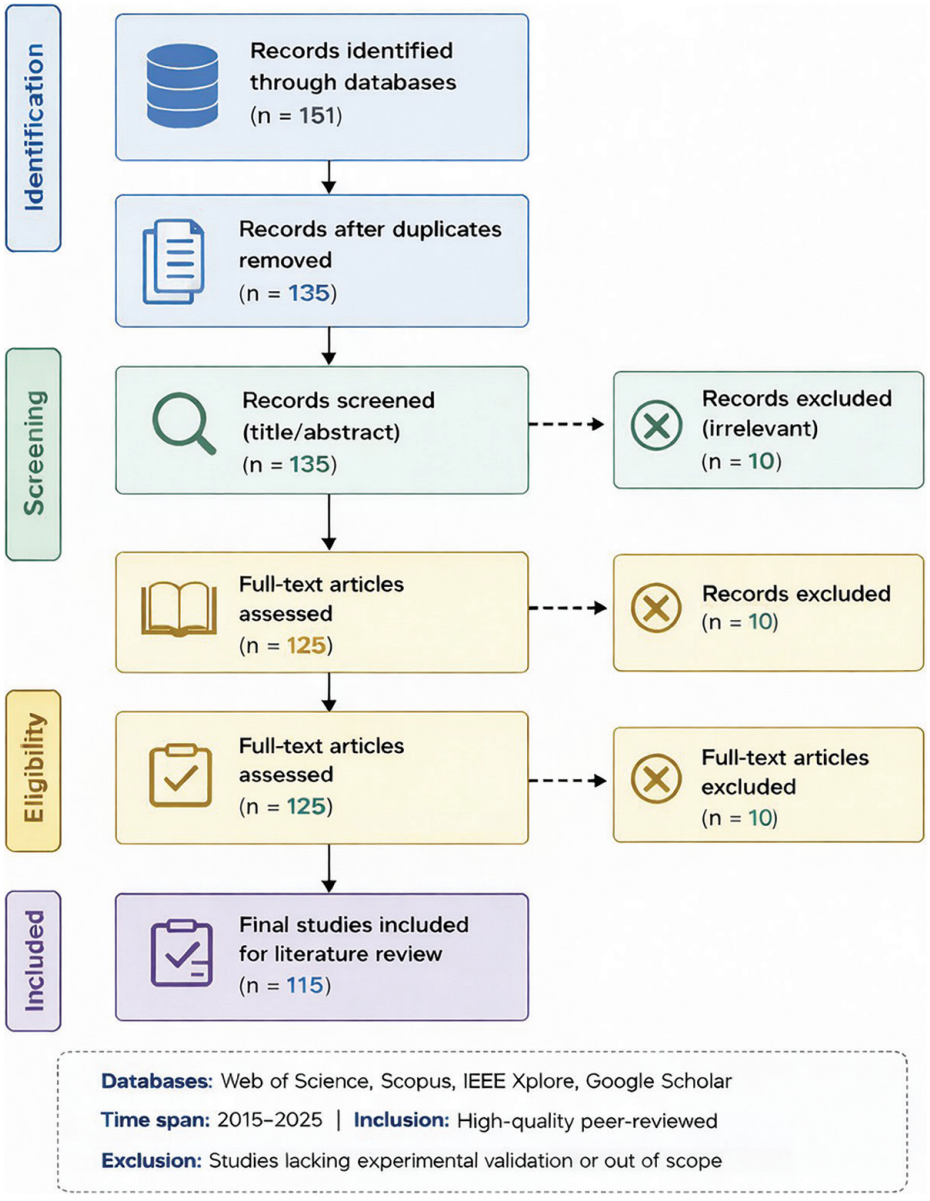


Figure 1: Flow diagram of systematic literature selection process.

analysis and identify emerging research trends. This multidimensional classification framework enabled a comprehensive synthesis of the current state of sensor-enabled bio-acoustic systems while supporting the identification of technological trends, methodological limitations, research gaps, and future opportunities discussed throughout this review.

Figure 1 presents the flowchart of the systematic literature selection process.

Figure 2 provides a visual depiction of the role of different speech signals in detecting the different conditions in dairy animals.

To provide a structured and comparative understanding of existing literature, the reviewed studies are systematically categorized based on acoustic feature representation, AI methodologies, dataset characteristics, sensing modalities, application domains, and deployment architectures. The taxonomy used in this study highlights the evolution of bioacoustic monitoring systems and identifies key research trends, technological gaps, and opportunities for intelligent real-time sensing frameworks.

Figure 3 presents an infographic of taxonomy used in AI-driven bioacoustic monitoring systems.

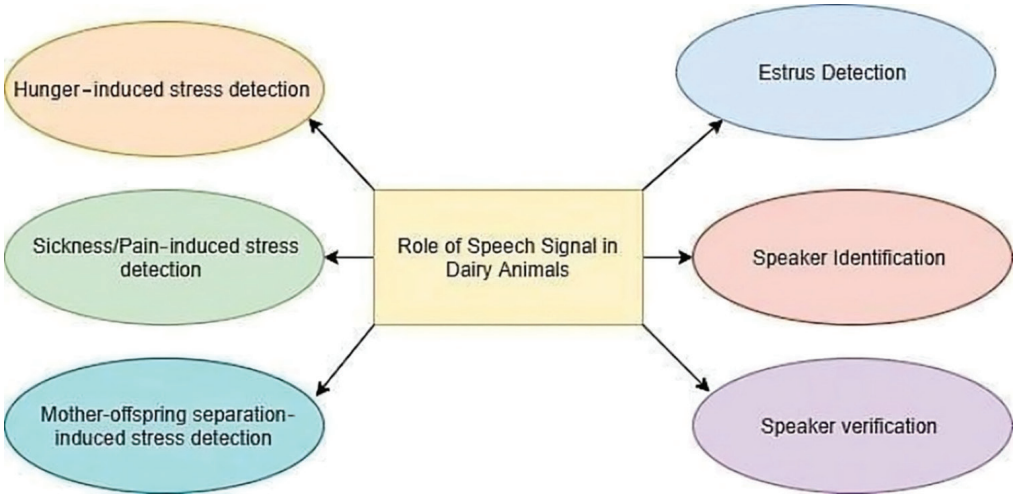


Figure 2: Role of vocal sounds representing different conditions in dairy animals.

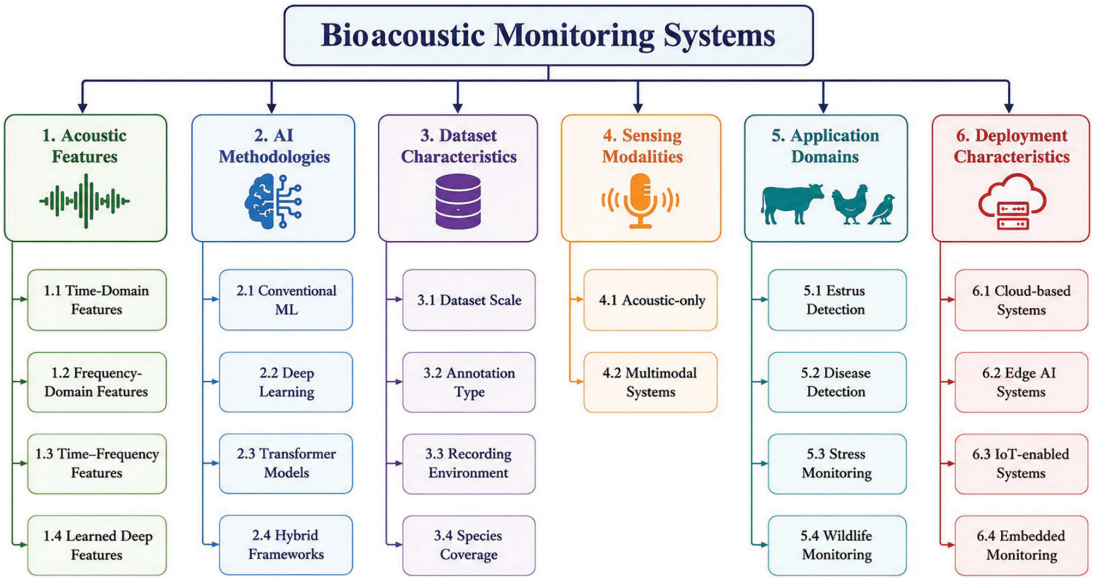


Figure 3: Taxonomy of AI-driven bioacoustic monitoring systems. AI, artificial intelligence; AI-driven, artificial intelligence-driven; ML, machine learning.

The organization of this paper is as follows: Section II provides the detailed review on animal welfare background and reviews existing studies related to stress detection, speaker identification, and estrus detection using bioacoustics, along with the application of neural networks in this field. Section III then discusses the major challenges and gaps identified in the current body of literature. Finally, Section IV presents the conclusion and future directions of the paper.

II. Detailed Review on Animal Welfare

Stress negatively affects dairy animal welfare and productivity, resulting in economic losses [4, 55]. Common stressors such as environmental changes, maternal separation, and diseases alter vocal and physiological responses [84, 85]. Increased vocalizations, including humming and semi-humming calls, serve as reliable stress indicators [45, 55]. By having

monitoring systems, identifying stress conditions and respiratory diseases could help improve the animal welfare. Using the ML-based audio monitoring systems, we can detect stress and respiratory diseases across various species with high accuracy [51, 58]. Speaker identification leverages unique acoustic features like MFCCs and pitch for individual animal recognition, facilitating herd management [54, 59]. Estrus detection studies correlate vocal features such as call duration and intensity with reproductive stages, improving breeding outcomes [16, 28, 72]. DL architectures, especially CNNs, enhance classification but face challenges including limited training datasets and false positives [37, 79]. Integrating physiological and acoustic data through IoT offers promise for real-time welfare monitoring [53, 86]. Despite progress, comprehensive vocal behavior research in dairy animals remains limited and requires further development [55]. Recent studies have deployed data analysis and computational methods where acoustic signals are transformed into time-frequency representations such as Mel-spectrograms and MFCCs, which serve as inputs to ML models. Classification performance is evaluated using metrics including accuracy, precision, recall, and F1-score. The current section presents a comprehensive overview of the studies conducted in areas, such as stress detection, speaker identification, estrus detection, and neural networks for bio-acoustic studies [5, 76, 87, 88].

Figure 4 presents the overview of vocalization analysis system architecture presented in this section.

a. Stress detection

Stress impedes the welfare of dairy animals, which further creates notable economic implications for the industry that demand attention. Usually, barn animals and chickens kept in captivity are exposed to diverse conditions of stress in farming and commercial poultry facilities. Few of the routine management practices, like separation from mother along with environmental pressures, induce stress in pigs, birds, bovines, etc., as shown in Table 1. Stress denotes the response of a living being to external stimuli present in the environment that it recognizes as a real or predicted threat to its well-being or survival. Stress-inducing elements in the environment, such as fear, temperature, and sickness like respiratory disorders, have a damaging impact on their productive performance.

The studies summarized in Table 1 reveal a clear trade-off between predictive performance and practical deployment. Behavioral and statistical approaches require minimal computational resources and are easy to implement, but they lack automation and scalability for continuous monitoring. Traditional ML methods improve automation while maintaining relatively low hardware requirements; however, their dependence on handcrafted features often limits robustness across different species, farms, and acoustic conditions. DL architectures consistently achieve higher classification accuracies by automatically learning complex stress-related vocal patterns, yet they typically require larger datasets, greater computational power, and increased memory resources. Furthermore, while recent CNN, LSTM, and transformer-based models show strong potential for large-scale stress detection, their suitability for real-time edge deployment remains insufficiently explored. Consequently, future research should focus on developing computationally efficient and noise-resilient frameworks that balance classification performance with hardware constraints, scalability, and operational requirements in commercial livestock environments.

The following section will explore the studies related to speaker-identification using animal vocalization.

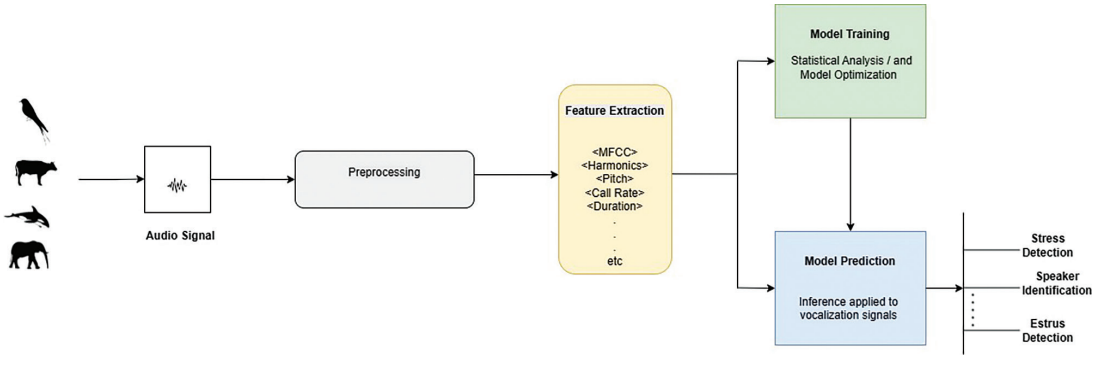


Figure 4: Vocalization analysis system architecture. MFCC, mel-frequency cepstral coefficients.

Table 1: An overview of multiple techniques used to detect stress in animals

Ref. No.	Dataset	Methodology	Findings	Observations
[88]	Thirty Holstein cows and their calves (a total of 60 animals), kept in separate barns, were monitored continuously for 10 hr using both audio and video recordings on the 1st, 6th, and 11th day after partial and full separation.	Video observation for physical signs and vocalization behavior was recorded. ANOVA and Tukey's test were conducted to study the effect on the frequency of vocal signals for calves in the case of partial separation as well as complete separation.	Separation significantly influenced calf vocalization frequency under both partial ($F_{2,28} = 8.59$, $p < 0.05$) and complete ($F_{2,28} = 17.26$, $p < 0.001$) conditions. During partial separation, calves vocalized more on days 1 and 6 (Tukey's test, $p < 0.05$) and showed a marked decrease by day 11 ($p < 0.01$). No significant changes were observed in cow vocalizations during partial separation ($F_{2,28} = -2.14$, $p = 0.27$). In contrast, cow vocalization frequency varied significantly during complete separation ($F_{2,28} = 19.35$, $p < 0.001$), with the highest levels on day 1 and reductions on days 6 and 11 ($p < 0.01$). Vocalization frequencies for both cows and calves did not differ on day 1 between partial and complete separation, though notable differences emerged between the two conditions across subsequent days.	The study showcased higher occurrence of vocalization levels and higher intensity, along with behavioral changes in calves and cows after partial separation. Immediate separation of cow and calf results in high stress conditions in both the cow and its calf.
[89]	The vocal recordings of 12 Holstein dairy calves (5 females and 7 males) 6 days post-birth, were relocated into an individual pen for calves having an area of 3.0m × 2.0m.	The vocalizations were classified into SHVs, HVs, and GVs. HVs consisted of low, continuous calls with a steady tone, whereas SHVs exhibited similar low, continuous calls with tonal variation. Categorization was based on waveform morphology and spectrogram patterns. Acoustic parameters were statistically evaluated using ANOVA within the GLM framework. Discriminant analysis with pairwise comparisons was employed for further differentiation. All parameters are reported as mean values with their corresponding standard deviations.	The intensity of HVs was the lowest at 64.9 ± 3.7 dB, while GVs exhibited the highest intensity at 75.2 dB ± 3.0 dB. The third formant was lowest for GVs (2242.0 Hz ± 57.4 Hz) and highest for HVs (2368.0 Hz ± 95.0 Hz). These findings indicate that fundamental frequency, duration, formant values, and intensity were effective parameters for distinguishing among vocalization types. Significant differences were observed across vocalization categories for these acoustic features ($p < 0.0001$). Except for the third formant and intensity, no other parameters differed between HVs and SHVs. Calf gender was included in the GLM analysis but showed no significant effect ($p > 0.05$).	The results claim that analysis of vocalization could be a suitable tool for evaluating the emotions of calves. SHVs and HVs are strongly related to separation or weaning stress.

(*Continued*)

Table 1: Continued

Ref. No.	Dataset	Methodology	Findings	Observations
[90]	The audio data of 120 Hy-line brown layer chickens (44 weeks old) were recorded.	The proposed automatic online monitoring system incorporates a pre-processing unit, a feature extraction module, a stress detection component, and a classifier. During pre-processing, real vocalizations of laying hens are captured through an audio sensor or CCTV camera. The detection and classification module identifies stress by analyzing variations in the acoustic signals and categorizes them into specific stress types, such as physical stress caused by temperature fluctuations and psychological stress resulting from fear. Finally, the system was evaluated experimentally using real audio recordings.	The system achieved a stress-detection accuracy of 96.2%, with an overall classification accuracy of 96.7%	Experiment performed on a limited dataset. No augmentation techniques were performed on the data. No techniques for data augmentation were performed.
[73]	Audio data of 20,000 white feather broilers was collected for a duration of 1 week (24/7) in a broiler house.	A broiler abnormal sound detection system was proposed for detecting respiratory disorders. A recognition algorithm was developed using WMFCCs, CDF, and HMM	An accuracy of 93.8%, along with a precision of 94.4%, recall of 94.1%, and an F1-score of 94.2%, was achieved in classifying stress levels in white-feather broilers	The results indicated that sound analysis can be effectively used to evaluate respiratory disorders in broilers within a commercial farm setting.
[12]	Audio data of pigs kept in a pigpen was acquired using audio sensors.	To analyze screams due to respiratory diseases and weaning pigs, the proposed system has the following assembly that makes it executable on the embedded board TX-2. It includes a module that collects audio signals, a noise-robust pre-processing module that detects sound regions in the signals and transforms them into spectrograms, and a pig anomaly detection module built on MnasNet coupled with an 8-bit filter clustering technique. The system, implemented on an embedded TX-2 board, integrates an audio signal acquisition module, a noise-robust preprocessing unit for detecting sound-active regions and generating spectrograms, and a pig anomaly detection module based on MnasNet with an 8-bit filter clustering technique for efficient and accurate detection.	An F1-score of 0.947 was achieved for abnormality detection, with consistent performance in noisy conditions, where the F1-score reached 0.94	The detection was performed in real time. The same approach can be used for cattle. The TX-2 board is suitable for edge deployment.

(*Continued*)

Table 1: Continued

Ref. No.	Dataset	Methodology	Findings	Observations
[91]	Cough sound data from 62 calves were continuously recorded over a 3-month period using a pig cough monitoring system (SoundTalks, Leuven, Belgium) to identify BRD. The setup included a C-4 small diaphragm condenser microphone (Behringer, Germany) and a MAYA44 sound card (ESI, Germany).	An algorithm was developed to analyze sounds and differentiate calf coughs from other vocalizations. It consisted of three stages: sound detection, feature extraction, and classification, and was implemented in MATLAB 8 (The MathWorks, Inc., USA) using 60 days of data from each housing unit. Throughout the recording period, a trained professional conducted weekly health assessments, and calves diagnosed with BRD received anti-inflammatory and antibiotic treatments, with treatment dates documented. These records were used to examine the relationship between cough frequency and BRD progression.	The proposed algorithm achieved sensitivity of 50.3%, specificity of 99.2%, and precision of 87.5% in the cough identification task. In four out of five periods, the surge in coughing frequency corresponded with the development of BRD in calves. The period of elevated frequency of coughing was always witnessed before the treatment of calves.	The ratio of cough calls was 6.3% that leads to low sensitivity but higher specificity. The proposed system has the potential to detect BRD in calves at the initial stage.
[92]	The study was conducted on two crossbred cattle herds located in separate fields. Herd 1 consisted of 21 adult multiparous females, while Herd 2 included 23 adult multiparous females, all maintained on a free-range farm i n Nottinghamshire, U.K.	Playback experiments were conducted, and binomial GLMMs were performed. The GLMM analyses were carried out using R v2.13.0 (R Development Core Team, 2009, R Foundation for Statistical Computing, Austria). For each model, the significance of the factors was assessed by comparing model fits.	The results indicated that vocal recognition between cattle mothers and their offspring operates in both directions. Mothers with younger calves showed a stronger response to playbacks of their own calf's calls than those with older calves, suggesting that maternal responsiveness is affected by the calf's age.	The responses of mothers to calf vocalizations are affected by the age of the calf. Only behavioral responses were recorded in response to playback calls.

(*Continued*)

Table 1: Continued

Ref. No.	Dataset	Methodology	Findings	Observations
[93]	Audio recordings of laying hens under control and under stressors (umbrella opening, dog barking), across different ages, pre- and post-stress exposure.	A CNN model using MFCC features to classify vocalizations (control vs stress), also differentiating treatment types, age, and exposure times.	The model achieved precision and recall both exceeding 0.99, indicating very high correctness in distinguishing stressed vs non-stressed vocalizations.	These near-perfect metrics might reflect relatively well-controlled conditions; potential overfitting or limited generalization may exist with real-farm noise & variability. Age and type of stressor strongly affect vocal responses.
[94]	Labeled vocal corpora of laying hens under stress and control, including different age categories and exposure conditions.	CNN model trained on MFCC spectral features; classification of stressor type, age, and exposure conditions.	The model achieved ~94% overall classification accuracy in discriminating stress vs non-stress, stressor types, etc.	The inclusion of age/exposure metadata improves robustness in stress identification of hens; spectral MFCC features are strong discriminators for stress vocal cues. The experiments are lab-based; external noise & field condition robustness need to be further tested.
[95]	Audio of 20 dairy cows under negative affective state (isolation); calls labeled as HF vs LF, also individual cow identity labels. Total ~1,144 records.	Two frameworks: an explainable ML model and a DL model. Tasks: classify HF vs LF calls, and identify individual cows from vocalizations.	For LF vs HF classification: 87.2% (explainable model) and 89.4% (DL model) accuracy was achieved. For individual identification: the accuracies were 68.9% (explainable) and 72.5% for explainable model and (DL model respectively).	Useful demonstration that vocal features carry both emotional (affective) and individual identity information. However, individual ID performance is modest; more data or better modeling might improve that.

(*Continued*)

Table 1: Continued

Ref. No.	Dataset	Methodology	Findings	Observations
[96]	Vocal recordings from broiler chickens exposed to environmental stressors (cold, heat, wind), with variants in dataset size, varying waveform lengths, and age input data.	ANN model with different parameter settings; experimented with adding age and longer waveform inputs; also used majority vote among models.	The best model achieved mAP = 0.97 on the larger dataset.	Very high performance suggests strong capacity to distinguish stress types; inclusion of additional metadata (age, etc.) helps; performance may depend on quality of recordings and noise levels.
[97]	Audio recordings of 14 water buffalo dams and 13 grey cattle dams collected on a Hungarian farm. Total ~139 calls recorded 2–12 hr post calf handling. Recordings done 2–3 m from animals, sample rate 32 kHz, capturing vocalizations under visual/ auditory separation from calves.	Acoustic signal processing: pitch estimation via MATLAB Audio Toolbox, smoothed with a sixth-degree polynomial. Spectral analysis, formant analysis, energy unevenness (H), MFCC computation (14 coefficients). Statistical tests: Mann–Whitney *U*, Welch's *t*-test.	Median pitch was significantly higher in gray cattle (6.579 $\log_2$(Hz) ≈ 95.6 Hz) compared to buffalo (5.887 $\log_2$(Hz) ≈ 59.2 Hz; $p \approx 7.52 \times 10^{-14}$). Buffalo showed a greater pitch change rate (0.820 octaves/s) than gray cattle (0.259 octaves/s; $p \approx 2.36 \times 10^{-10}$). Energy unevenness (H) was lower in gray cattle (1.895) than in buffalo (2.176; $p \approx 0.001356$). Significant interspecies differences were observed in MFCC values—MFCC4 (grey cattle: 0.536; buffalo: –0.910; $p \approx 2.2 \times 10^{-16}$) and MFCC8 (grey cattle: 0.168; buffalo: –0.389; $p \approx 5.9 \times 10^{-10}$). No significant within-species individual variation was detected.	Species-specific acoustic signatures were detected after separation stress, useful for monitoring dam emotional arousal. MFCC features and pitch dynamics are reliable discriminators. Buffalo calls are more dynamic, potentially indicating higher emotional arousal. Small sample size and farm-specific conditions limit generalization. Potential application in precision livestock farming, automated welfare assessment, and bio-acoustic stress monitoring.

(*Continued*)

Table 1: Continued

Ref. No.	Dataset	Methodology	Findings	Observations
[62]	Audio recordings of 3,302 calls from captive coppery titi monkeys (*Plecturocebus cupreus*) across three zoo populations. Calls manually labeled into 10 types (3 short calls, 5 long calls, 2 harsh calls).	LSTM neural network for call detection and classification. Data preprocessing included segmentation, feature extraction, and normalization; model trained and validated on labeled datasets.	Call detection accuracy: 95% Call classification accuracy: 97%	Demonstrates species-specific vocal patterns can be reliably detected using ML. Useful for welfare monitoring in captivity and for conservation of endangered species. Success despite small and unbalanced dataset shows LSTM's effectiveness for wildlife bio-acoustics. Could be adapted for automated monitoring in field or captive environments.

ANN, artificial neural network; BRD, bovine respiratory disorders; CDF, correlation distance fisher criterion; CNN, convolutional neural network; DL, deep learning; GLM, generalized linear model; GLMMs, generalized linear mixed models; GVs, general vocalizations; HF, high-frequency; HMM, hidden Markov model; HVs, humming vocalizations;, LF, low-frequency; LSTM, long short-term memory; mAP, mean average precision; MFCC, mel-frequency cepstral coefficients; ML, machine learning; SHVs, semi-humming vocalizations; WMFCCs, wavelet transform mel frequency cepstrum coefficients.

b. Speaker identification

Voice-recognition systems have become an important application of speech recognition technology [69]. In animals, these systems aim to assist humans in distinguishing and identifying which individual is vocalizing [69]. Because different animals produce characteristic acoustic frequency patterns, the accuracy of detecting an individual's unique ID is comparatively high. Such systems are especially valuable in large herds or populations, including zoos, farms, and veterinary hospitals [69]. Research on animal identification using vocal pattern recognition algorithms remains limited [76]. Ongoing initiatives involve collecting vocalizations from species such as tamarins, beluga whales, dolphins, and various birds. Although each species exhibits distinct vocal traits that add analytical complexity, existing systems provide a flexible foundational framework for developing similar technologies for other animals [87]. Further advancements plan to incorporate additional bioacoustic features, including vocalization duration, bandwidth, and fundamental frequency [33, 69, 76, 87]. The related work is summarized in Table 2.

The studies summarized in Table 2 demonstrate that speaker identification performance is closely linked to model complexity and data availability. Classical approaches based on MFCCs, linear discriminant analysis (LDA), DTW, and HMMs offer low computational overhead and modest hardware requirements, making them attractive for resource-constrained applications. However, their performance tends to decline when vocal complexity, individual variability, and species diversity increase. In contrast, approaches incorporating temporal modeling and richer acoustic representations generally improve recognition accuracy but require larger datasets and greater computational resources. Furthermore, most reported systems have been evaluated on relatively small and controlled datasets, leaving questions regarding scalability and robustness in large-scale monitoring scenarios unanswered. Although lightweight feature-based methods are more suitable for edge deployment, advanced learning frameworks are likely to provide superior recognition capability at the cost of increased processing and memory demands. These observations suggest that future speaker-identification systems should balance recognition performance with computational efficiency, scalability, and real-time deployment requirements to enable reliable operation in practical livestock and wildlife monitoring environments.

The upcoming section presents studies related to estrus condition detection in animals using animal vocalization.

c. Estrus detection

There is a high economic importance of estrus detection in cattle [27, 28]. Accurate detection of the estrus stage in cattle is necessary to identify the appropriate time of insemination [2, 27, 28]. Incorrect and inaccurate heat detection is a major constraint in the reproductive performance of cattle herds [2, 3, 16, 27–31]. In an organized dairy herd, about 50% of the heats are missed due to no detection, and 20.75% of buffaloes are inseminated at the incorrect time [2, 28]. Each missed heat leads to the loss of a complete estrus cycle (21 days) that in a seasonally calving herd equates to a loss of 21 days of potential milk production; hence, each missed heat has a notable financial loss [2, 27]. Limited research work has been conducted in order to develop a user-friendly non-invasive system of estrus detection [2, 16, 28, 30, 31]. Attaining higher estrus detection efficiency and accuracy is the principal key to improving reproduction management in cattle [2, 16, 28, 29, 31]. The related work has been presented in Table 3.

The evolution of vocalization-based estrus detection has progressed from behavioral observation and statistical analysis to ML and multi-modal DL frameworks.

The studies presented in Table 3 reveal a clear trade-off between detection performance, implementation complexity, and deployment practicality. Behavioral observation and threshold-based approaches require minimal computational resources and can be implemented with relatively simple sensing infrastructure; however, they often suffer from subjectivity, limited scalability, and reduced robustness under variable farm conditions. ML methods based on handcrafted acoustic features improve automation while maintaining modest hardware requirements, making them suitable for low-cost deployments. Nevertheless, their performance is highly dependent on feature quality and may deteriorate in noisy environments. More recent DL and multi-modal architectures achieve superior estrus-detection accuracy by learning complex acoustic and physiological patterns directly from data, but these gains come at the cost of increased computational, memory, and sensor requirements. In particular, multi-modal systems require additional sensing hardware and data synchronization mechanisms, which may limit adoption in resource-constrained farming environments.

Table 2: An overview of approaches used in speaker identification

Ref. No.	Dataset	Methodology	Findings	Observations
[69]	Two audio datasets of 19 types of cricket calls and 30 types of frog calls were analyzed.	The acoustic signal was assumed as a sequence of syllables. The syllables were segmented before the recognition process. The proposed method was based on averaged MFCCs and LDA to automatically identify animals using their vocalizations. The proposed method was employed to reduce feature dimensionality and improve classification accuracy.	The average classification accuracy was recorded as 98.1% and 96.8% for 19 types of cricket calls and 30 types of frog calls, respectively.	This simple approach is effective for identifying crickets and frogs, as their sounds are less complex than those of other animals. However, it remains uncertain whether this method can be applied to more complex vocalizations, such as bird songs, which exhibit extensive variation in their syntactic structures.
[33]	The study was conducted on 250 voice samples of 25 lactating Karan Fries cows that were collected after parting of animals from the herd during the morning period	The vocalizations of individual cows were analyzed to extract acoustic features using PRAAT software (Paul Boersma & David Weenink, University of Amsterdam, Netherlands). It was followed by least square analysis to inspect the substantial differences in the vocal sounds.	The extracted acoustic features demonstrated that differences in amplitude, pitch, formants, intensity, pulse, period, voice breaks, jitter, shimmer, mean values, unvoiced frames, noise-to-harmonic ratio, total energy, and mean harmonic-to-noise ratio were highly significant ($p < 0.001$). Among these, only a subset of features—pitch, pulse, jitter, shimmer, voice breaks, formants, and HNR—showed significant differences ($p < 0.05$) between individual KF cows. Formant frequency emerged as the most distinctive acoustic feature, capable of differentiating individual cows based on their vocalizations.	The study can be extended to other cattle species as well.

(*Continued*)

Table 2: Continued

Ref. No.	Dataset	Methodology	Findings	Observations
[76]	The vocalization dataset of African elephants consisted of one 18-year-old male and six adult nulliparous females aged 19–30 years, housed and recorded at Disney's Animal Kingdom, Orlando, Florida, comprising a total of 74 vocalizations.	The vocalizations were transformed into a set of information-rich spectral features. For classification, 12 MFCCs were extracted, computed using a chirp Z-transform to interpolate the lower frequency range. The HMM technique was applied to capture the individual characteristics of each vocalization type. To increase the effective size of the training set, leave-one-out cross-validation was employed for testing. The HMM model was trained using the Baum-Welch algorithm, and the Viterbi algorithm was applied to test examples to determine the most probable sequence of states, indicating the duration of each state within the vocalization.	The vocalization-type classification experiment achieved an accuracy of 77.0%, while the speaker identification experiment attained an accuracy of 72.2%.	More work can be done towards achieving better accuracy. In comparison with the humans, the elephants vocalize at much lower fundamental frequencies. This frame-work can also be extended to other species having much lower fundamental frequencies than humans as well.
[87]	The dog bark calls audio dataset was used.	The proposed animal voice recognition system utilized ZCR, MFCC, and DTW algorithms to identify individual animal vocalizations.	The distance between the input sound and dog3.wav was 0.563745 at its maximum, whereas the distance to dog2.wav was 0.00 at its minimum. This indicates that the input sound corresponds to dog2.wav in the reference dataset.	No information about size of Dataset has been provided. No information about the accuracy of the proposed approach has been mentioned. Details regarding the dataset size and the accuracy of the proposed approach have not been provided by authors.
[98]	7,285 coo call-type vocalizations from 8 Rhesus macaques (4 males, 4 females).	Feature extraction involved MFCCs along with delta and delta-delta coefficients. The frame size was set to 4 ms with a step size of 2 ms. The model employed was a 4-state left-to-right HMM.	Gender recognition achieved an accuracy of 84.45% (1,233 correct out of 1,460). For speaker identification, males were correctly identified 91.08% of the time (633/695), females 83.27% (637/765), resulting in an overall accuracy of 81.85% (1,190/1,460 correct identifications).	Demonstrates the feasibility of using HMMs for automated gender recognition and speaker identification in non-human primates. Highlights the potential for applying this method-ology to other mammalian species for vocalization analysis. Provides a foundation for future studies on primate communication and bio-acoustic monitoring.

DTW, dynamic time warping; HMM, hidden Markov model; HNR, harmonic noise ratio; MFCCs, mel-frequency cepstral coefficients; LDA, linear discriminant analysis; ZCR, zero-crossing rate.

Table 3: An overview of techniques deployed in the area of estrus detection

Ref. No.	Dataset	Methodology	Findings	Observations
[27]	Voice samples of 30 buffaloes were collected between 6 AM and 8 PM during their four phases of the estrus cycle.	A DSS was developed for estrus detection in the Murrah breed, based on 15 identified acoustic features. The WEKA 3.6.2 (University of Waikato, New Zealand) platform was used to explore the classification procedure space. The 15 MFCCs were input into the C4.5 algorithm to generate classifiers, whose performance was validated against perceptual tests. Feature evaluation was conducted using the J48 classification method.	The accuracy and efficiency of the DSS model were 95% and 78.94%, respectively. Sensitivity was found to be 78.94%, and Specificity was achieved as 97.32%. Precision rate (true positive rate) was 81.08%. The false positive rate was calculated as 18.91%. The false negative rate was calculated as 3.04%.	Sensitivity and precision require improvement, and the false positive rate should be reduced. Additionally, recordings were not made under natural conditions, where ambient noise could interfere with the sound signals.
[28]	Voice recordings were obtained from 30 buffaloes between 6:00 and 8:00 AM across the four phases of the estrus cycle: estrus, diestrus, proestrus, and metestrus.	Acoustic features were extracted using the PRAAT 5.1.36 software, focusing on MFCCs. Threshold or cut-off values were calculated, and the acoustic data were analyzed using the least squares method. Duncan's Multiple Range Test was applied to determine the significance of differences among the various subclasses.	Significant differences ($p < 0.05$) were observed during estrus for call duration, intensity, formants 1–5, pitch, and noise-to-harmonic ratio compared to other phases. Estrus thresholds—call duration 2.17 s, intensity 61.41–81.99 dB, pitch 130.66–177.58 Hz, noise-to-harmonic ratio 7.47%, and formants 1–5 at 1281.28–6648.73 Hz—enabled 95% accuracy in distinguishing estrus from non-estrus vocalizations.	Vocalizations occurred irregularly, preventing measurement of inter-call intervals. The established thresholds of key acoustic features for the four estrus phases can serve as cut-off values in a DSS, generating alerts to assist farmers in timely insemination of buffaloes.
[32]	Dataset with 100 estrus sound samples and 180 normal sound samples of Korean native cows (Bos taurus coreana) was used.	A data-mining solution for estrus detection is proposed. The MFCCs were extracted from the dataset data followed by CFS for attribute subset selection. Finally, SSVD (support vector data description) was utilized for early anomaly detection.	Using 80% of randomly selected sound samples for training, the system achieved an average detection accuracy of 96.86%, with a false positive rate of 7.46% and an average false negative rate of 3.14%.	The sound data that was collected was only of two types: estrus and non-estrus. A limited and small dataset was used.

(*Continued*)

Table 3: Continued

Ref. No.	Dataset	Methodology	Findings	Observations
[66]	Sound recordings (Continuous in nature) of 10 German Holstein heifers in the periestrous phase were collected.	A recording algorithm deploying Graphic programming surface language LabVIEW (LabVIEW, 1999) was implemented along with envelopes of the cepstrums. The number of calls per hour was recorded. The calls of the heifers were classified into “moo” calls (harmonic sound) and “bellow” calls (noisy and non-harmonic). For vocalization rate, the rate of the bellow and moo calls was recorded. The Holstein heifers were parallelly administered concentrations of blood progesterone. The visual observation of estrus behavior was also examined.	The recorded sounds exhibited an augmented vocalization rate and an aggravated proportion of bellow calls during estrus phase. The increase in vocalization rate corresponded to visual signs of estrus observed. A steep elevation in vocalization rate was recorded 1 day prior to when estrus was detected. On day 0, the rate elevated again from 70 to 171 calls. The rate of vocalization increased by nearly 84% from day 2 to 0, and 59% from day 1 to 0. After day 0, the rate of vocalization dropped by about 79%. A similar pattern was detected for bellow and moo calls, but the number of bellow calls rose higher than the number of moo calls from day 2. There was an increment in parts of bellow calls on day 2 and a decrement from day 0 to 1. No correlation was found between progesterone and vocalization rate.	The study included only tethered heifers, and it remains unclear whether the vocalizations were induced by social stress or were associated with estrus.
[31]	Twelve German Holstein heifers were studied.	Estrus behavior, estrus scores, and vocalization data were analyzed using ANOVA with the MIXED procedure (Mixed Linear Models). LSM and standard errors were calculated, and all pairwise differences were assessed using the Tukey–Kramer test. The SLICE statement of the MIXED procedure was employed for partitioned anal-yses of LSM, allowing focused comparisons within factor levels. The results were subsequently compared with observed visual estrus signs.	During natural estrus, the rate of vocalization increased considerably toward reaching the climax of estrus behavior. This elevation was less evident during superovulation treatment.	The temporal association between vocalization rate and estrus behavior was revealed. Vocalization rate can be used as a prospective indicator of the peak of estrus in automated devices for estrus detection. Detection of individual vocalizations in large group-housed dairy animals is a big challenge.

(*Continued*)

Table 3: Continued

Ref. No.	Dataset	Methodology	Findings	Observations
[30]	Ten Holstein cows were taken into observation from a herd size of 150.	The study examined vocalization behavior during the estrus period. Statistical analyses were performed using SPSS 24.0. Descriptive statistics, including mean, median, frequency, standard deviation, and ratios, were calculated, and chi-square tests were applied to compare qualitative data. Results were reported at a 95% confidence interval with a significance level of $p < 0.05$.	During the estrus phase, vocalizations increased from 12.67 ± 2.11 to 61.46 ± 12.18 calls. In the post-estrus phase, vocalizations decreased to 24.00 ± 2.01. Additionally, the number of sound repetitions was significantly higher in cows in estrus compared to others ($p < 0.01$). Detection error rates ranged from 17% to 55%.	During the estrus period, cows exhibited increased calling activity. Considerable standard errors were observed in vocalization counts, reflecting a high rate of false alerts. Additionally, vocalization frequency showed substantial individual variability among the cows.
[29]	The voice recordings of Five group-, housed heifers each in the periestrus or estrus period were taken for 5 days. 1,404 vocalizations from the focus heifers and 721 vocalizations from group mates were obtained.	A collar-based CCM was developed, incorporating ABM and SBM sensors along with a recording unit. The system used a post-processing algorithm to identify the caller by mapping data collected from two microphones. Heifers wore the CCM for 5 days, and the recordings were analyzed in comparison with audio-visual data. Audio was sampled via the stereo line-in connected to the CCM's processing unit, with one channel assigned to the ABM and the other to the SBM. The microphone data were sampled at 22 kHz and stored on a microSD card. To qualify as a potential call and be saved, audio events had to meet three criteria: (1) the ABM signal frequency range of 50–2000 Hz, (2) ABM event duration between 0.5 s and 10 s, and (3) an SBM signal with an average of more than five zero crossings per window. Collars were removed daily to transfer data to a central storage location.	The results indicated that the CCM achieved a sensitivity of 87% and a specificity of 94%, with a negative predictive value of 80% and a positive predictive value of 96%.	The results indicated two types of data gaps: interruptions occurring within files and gaps between separate files. These issues reduce data continuity and accuracy, highlighting the need for technical improvements. Additionally, the device's design requires enhancement to ensure it can be comfortably and securely attached to the cattle's neck for reliable long-term use.

(*Continued*)

Table 3: Continued

Ref. No.	Dataset	Methodology	Findings	Observations
[99]	The study was carried out on 94 Murrah buffaloes at LRC of NDRI, Karnal, Haryana for a period of 4 months (November 2011–February 2012).	The present study aimed to investigate estrus-related behavioral signs in Murrah buffaloes of both primi-parous and pluri-parous parities. Visual observation of behavioral signs of estrus like bellowing and vocalization were used for assessing the onset and intensity of estrus were recorded. Statistical analysis was performed using the Chi-square test on SigmaPlot 11® software (Systat software Inc. USA) to calculate the difference between heifers, primi-parous and pluri-parous buffaloes for behavioral signs of estrus.	In addition to other behavioral indicators, the study found that bellowing/vocalization was significantly higher ($p < 0.05$) in primiparous buffaloes (77.27%) and pluriparous buffaloes (72.88%) compared to heifers (38.46%).	The vocalization of the buffaloes can be used for enstrus detection. The intensity of vocalization/bellowing, tail raising, and micturition behaviors vary greatly among different parities of Murrah buffaloes during estrus.
[100]	Audio recordings were obtained from riverine buffaloes under different physiological conditions, including estrus, delayed milking, and isolation. For each condition, a total of 600 acoustic samples were collected, comprising 300 samples confirming the condition and 300 samples not confirming it.	Acoustic features, including amplitude (P), total energy (P2s), pitch (Hz), intensity (dB), formants (Hz), number of pulses, number of periods, mean period (s), and percentage of unvoiced frames, were extracted using the MFCC technique. Predictive models were developed by training Decision Tree, Logistic Regression, and ANN models on partitioned acoustic data, with training-to-validation ratios of 60:40, 70:30, and 80:20. Decision Tree models were optimized using decision and average squared error (probability) criteria.	Decision tree models predicted estrus, isolation, and delayed milking with an accuracy of 66.1%, 84.3%, and 71.3%, respectively, while the logistic regression model predicted with an accuracy rate of 59.5%, 71.1%, and 65.7%, respectively, and the ANN model predicted these three conditions with 77.7%, 85.2%, and 79.4% accuracy, respectively.	The ANN model demonstrated the best performance based on the lowest misclassification rate using an 80%–20% data split. However, Decision Tree algorithms provided additional insights, indicating that maximum intensity, minimum amplitude, and the first formant (F1) are the most critical vocal features for identifying physiological conditions such as estrus, isolation, and delayed milking in dairy buffaloes.

(*Continued*)

Table 3: Continued

Ref. No.	Dataset	Methodology	Findings	Observations
[101]	Audio recordings and thermal infrared images of the vulvar and hip regions were collected from 59 Danish Landrace breeding pigs, including 19 in estrus and 40 in non-estrus. In total, 480 thermal images and 480 audio samples were gathered, evenly distributed between estrus and non-estrus conditions.	An Adaptive-PIG-OESTUS-CNN-ViT model was proposed, combining Vision Transformer and CNN architectures. The model extracted and fused features from multi-modal data, employing an adaptive cross-attention mechanism to automatically learn feature vectors representing both thermal infrared and audio inputs. These features were then processed by an enhanced DenseNet network to classify estrus and non-estrus states in breeding pigs.	The model attained an accuracy of 98.92%, a recall of 95.83%, and an F1-score of 97.35%, demonstrating effective non-invasive estrus detection in breeding pigs.	Compared to conventional estrus detection methods, this approach more effectively integrates multi-modal data to identify the estrus state in breeding pigs, offering an efficient, objective, and non-invasive method for sow estrus detection.

ABM, animal-borne microphone; ANN, artificial neural network; ANOVA, analysis of variance; CCM, cattle call monitor; CFS, correlation-based feature selection; CNN, convolutional neural network; DSS, decision support system; LRC, livestock research centre; LSM, least squares means; MFCC, mel-frequency cepstral coefficient.

Consequently, future re-search should focus on scalable and computationally efficient estrus-detection frameworks that balance predictive accuracy, robustness, hardware cost, and real-time edge deployment feasibility across diverse livestock production systems.

The next section highlights the studies related to neural network architectures employed in the area of bioacoustics.

d. Neural networks in bioacoustics

In recent times, a few studies have employed DL approaches in the area of bioacoustics as well. Jung et al. [91] and Nanni et al. [102] have conducted studies on sounds of various species. One of the key challenges in deploying ML techniques in the animal sound dataset is the limited size. The high rate of false positives is another concern that needs to be addressed using suitable techniques [88–92, 102]. The authors implemented CNN models on the animal sound corpus. In a few studies, acoustic signals are transformed into time-frequency representations such as Mel-spectrograms and MFCCs, which serve as inputs to ML models. Classification performance is evaluated using metrics including accuracy, precision, recall, and F1-score. However, a few data analysis and computational studies have assumed signal stationarity and balanced datasets, which may not apply in real-world low-resource datasets due to limitations in generalizability. In most of the related work, the system architecture deployed typically involves sensor nodes that capture acoustic signals, which are then transmitted to edge or cloud processing units. Latency and synchronization issues arise due to asynchronous sensor inputs and network delays. These have been mitigated through buffering strategies, timestamp alignment, and edge-level preprocessing. The related work has been presented in Table 4.

The studies summarized in Table 4 illustrate a clear trade-off between predictive performance and deployment efficiency. CNN-based models offer a favorable balance between accuracy and computational cost, making them suitable for real-time monitoring and edge-based bio-acoustic applications with limited hardware resources. The incorporation of transfer learning and data augmentation further improves performance without substantially increasing deployment complexity. In contrast, Transformer-based and hybrid architectures achieve superior recognition capability by modeling long-range temporal and contextual relationships; however,

Table 4: An overview of related work in the area of application of neural networks in bioacoustics

Ref. No.	Dataset	Methodology	Findings	Observations
[91]	Audio recordings from 130 breeding Korean native cows, including 30 calves aged 12–40 months, 30 calves aged 6–8 months, and 50 calves below 6 months.	A web-based monitoring and recording system using AI for cattle sound classification was developed. Twelve sound sensors captured audio in real time. Voice data were converted into MFCC features and input into a CNN model. STFT was applied for noise reduction. Audio was labeled into four classes: estrus call, cough, feed anticipation call, and normal call.	The CNN model achieved 91.38% accuracy in cattle sound recognition, which improved to 94.18% after noise filtering. A total of 897 classification records were used to build the final model, achieving 81.96% overall accuracy. Accuracy for estrus calls, coughing, feed-based sounds, and normal calls was 95.4%, 86%, 74%, and 76%, respectively.	Misclassifications occurred mainly with feed related and normal sounds being identified as estrus calls (18% and 14%, respectively). The model can be further optimized to improve accuracy.
[103]	Custom field recordings of chicken vocalizations taken in commercial farm settings. Total: ~12,500 labeled audio samples (healthy vs. Newcastle disease affected). Sample duration: 1–5 s. Data includes background farm noise; balanced classes. Labels confirmed by farm diagnostics and manual annotation.	Developed Deep Poultry Vocalization Network (DPVN), a custom CNN architecture. Input: Mel-spectrogram representation of each audio clip. Training was monitored using accuracy and loss functions over multiple epochs. Early stopping was applied to prevent overfitting.	Accuracy achieved is 97.2%, value of sensitivity achieved is 96.8%, value of specificity attained is 97.5%	Demonstrated robust Newcastle disease detection under real farm noise, suitable for non-invasive early warning.
[104]	Aggregates~1,000–20,000 labeled samples per study (images and sound), focused on common poultry diseases (Newcastle, Salmonella, Coccidiosis); collected via IoT, smart-phones, farm sensors	Surveys CNN, RNN, YOLO, and hybrid deep models; emphasizes end-to-end pipelines and data augmentation for robust detection.	DL models achieve 85%–98% accuracy depending on input/data; image-based CNN approaches typically report >90% accuracy for well-represented classes.	DL methods far outperform manual approaches, but rare diseases need better data; recommends mobile deployment and more di-verse dataset collection.
[105]	Recordings from 74 dogs, spanning multiple breeds, ages, and sexes; ~4,000 annotated bark/howl samples recorded in varied contexts (lab and home environments)	Feature extraction using STFT, MFCC, LFCC, fed into CNN classifiers including AlexNet, DenseNet, EfficientNet, ResNet50, ResNet152; evaluated classification performance for different vocal behavior types	Achieved high classification accuracy (~94%) and F1 score (~0.92) with CNN models; pretraining on large human speech datasets (Wav2Vec2) improved representations for dog bark tasks, outperforming baseline models up to ~70% accuracy in specific classification tasks.	Demonstrated the feasibility of fully automated dog vocalization analysis; pre-trained human speech models benefit animal vocal task performance; introduced a publicly available dataset.

(*Continued*)

Table 4: Continued

Ref. No.	Dataset	Methodology	Findings	Observations
[43]	Small mammal calls, 7 species, 100 calls/class (total ~600), field and captive	Transfer learning with pre-trained CNNs fine-tuned on mammal vocalizations; heavy data augmentation.	Accuracy ranged between 78% and 85% across species and call types.	Demonstrated feasibility of DL classification on small, imbalanced bioacoustic datasets with augmentation.
[106]	Chicken vocalizations (feed, stress, laying); ~5,400 clips recorded on-device, various noise levels; all classes >1,000 clips	TinyML neural networks optimized for edge processors, CNN-based architectures, real-time processing.	Achieved accuracy of 92.4%, with low latency (<5 ms), enabling on-farm real-time monitoring.	Demonstrated viability of resource-constrained AI models for continuous vocal welfare monitoring in poultry farms.
[63]	Curated dataset of 17,653 bird audio clips (3 s each) from 18 species, balanced, recorded across varied natural environments.	Two DL models based on Audio Spectrogram Transformer architecture; one for detecting animal sounds (precision 98.3%, recall 99.3%), second for classifying wolf howls (precision 89.6%, recall 93.4%).	Precision & recall for detecting animal sounds is 98.3% & 99.3% respectively. Precision & recall for classification of wolf howls are 89.6% & 93.4%, respectively.	Models enable efficient, automatic detection of wolf howls, crucial for large-scale ecological and conservation monitoring.
[107]	Vocalizations of 3 male Bengalese finches were recorded for 30 days, 35 days, and 105 days, yielding 13,288, 4,492, and 14,435 songs respectively. These were converted to sequences of discrete units.	FinchGPT: Transformer-based language model (similar to GPT-2) trained to capture long-range dependencies in birdsong syntax using sequence tokenization and self-attention mechanisms.	Outperformed RNNs, Markov models in sequence prediction and classification; attention analysis reveals key long-range dependencies aligning with biological song structure.	Demonstrates strong potential for deep language models in animal communication; integrates semantic and syntactic information; opens new avenues for understanding animal vocal syntax.
[108]	Elephant calls, open field, ~7,000 labeled vocalizations, five sub-types, background inclusive	Deep CNN models optimized through hyperparameter tuning for sound classification.	Achieved accuracy of 91.5% and recall of 90%.	Optimization strategies contribute to improved classification performance, aiding conservation efforts.
[50]	Himalayan Monal: Custom dataset (~350 calls, 21 individuals), background noise from Himalayan field recordings.	Evaluated CNN, VGG, ResNet architectures for bird call classification; experimented with transfer learning and data augmentation strategies.	Accuracy achieved ranged between 87% and 90%; achieved precision of 88%.	Proved DL useful-ness for rare, elusive bird monitoring where data is limited; recommended for conservation applications.

(*Continued*)

Table 4: Continued

Ref. No.	Dataset	Methodology	Findings	Observations
[109]	Curated dataset of 17,653 bird audio clips (3 s each) from 18 species, balanced, recorded across varied natural environments.	Proposed DuSAFNet: multi-path feature fusion network using spectral-temporal attention mechanisms to leverage both frequency and temporal cues.	Accuracy attained is 96.88%, F1-score value is 96.83% on bird audio classification.	Effectively managed do-main shifts across recording sites; attention enhanced species-level recognition for practical field deployment.
[110]	Camera trap: 11,800 labeled audio clips, multiple South African species, paired with video, field conditions	DL ensemble combining CNNs for image and bioacoustic data fusion.	Accuracy achieved for species ID is 89%. Achieved counting precision of 91%.	Multi-modal fusion enhances wildlife monitoring accuracy in natural habitats.
[111]	Over 45 hr of synchronized audio-video recordings from 18 sows in indoor housing; 1,236 annotated nursing behavior events including calling and non-calling contexts	Proposed TMF combining unimodal self-attention, audio-visual interaction enhancement, and adaptive dynamic decision fusion modules. Used sound source localization for sound direction.	Attained accuracy of 98.42% for general nursing behavior; 94.37% for fine-grained nursing behavior types.	Multi-modal fusion significantly improved recognition accuracy over unimodal; fine-grained classification critical for monitoring sow health and management in complex farm environments.
[64]	Newborn giant panda calls: ~600 clips (high-pitch, short duration), zoo institution, newborn phase	Proposed hybrid model combining LSTM with local-global parallel Trans-former network for capturing temporal patterns in neonatal calls.	Achieved recognition accuracy of 93.2% and F1-score of 91.8%.	Hybrid architecture improves neonatal call recognition performance compared to standalone LSTM or transformer models.
[44]	Marine mammal call sets: 11,400 calls (whales, dolphins), 6 datasets, different devices, variable SNR	MT-Resformer, a multi-scale, two-channel fusion network combining spectral and temporal information with transformer blocks.	Achieved accuracy of 96.1%, F1-score of 95.4%; outperformed other CNN and RNN baselines by 5%– 7%.	Demonstrated robustness to noisy, low-SNR underwater environments, effective for bio-acoustic monitoring of marine mammals.
[112]	Four geographic datasets of humpback whale song units; 16,800 recordings, cross-region, varying background (boat, surf)	CNN with MSDFA to capture call variability across regions; spectrogram-based inputs; trained for cross-region generalization.	Achieved classification accuracy of 94.1%, F1 score of 93.6%; MSDFA outperformed baseline CNNs by 3%–5%.	Approach robust to regional vocal variability and noise, improving conservation monitoring capabilities.

(*Continued*)

Table 4: Continued

Ref. No.	Dataset	Methodology	Findings	Observations
[113]	Swine barn ambient audio, 17 months, 12,800 sound clips; labeled as "event/non-event" against manual logs.	CNN-based audio event detection model trained with weak labels and data augmentation to handle noisy barn environment.	Achieved an F1-score of 83.5%, accuracy of 85%; performed well despite limited, weakly labeled data.	Showed feasibility of non-invasive, large-scale welfare monitoring in intensive farming environments despite acoustic complexity.
[61]	African elephant rumbles and sub-calls; ~10,300 audio segments, collared in wild and facility	CNN-based detector-classifier pipeline with endpoint detection; tackled calls and sub-call structure	Detection rate achieved is 92.7%; achieved sub-call classification F1 as 88.5%.	Automated endpointing is critical for call segmentation; the model integrates well with bio-acoustic monitoring systems.
[114]	Underwater: ~9,250 audio samples of target/non-target, 14 classes, 1–10 s duration, varying depths	Proposed two-stream network combining temporal and frequency components via transformer attention mechanisms.	Achieved accuracy as 94.6%, F1 as 94.2%.	Multi-stream fusion leverages complementary features for robust under-water target recognition.
[94]	Large-scale IoT-collected dataset of more than 250,000 chicken vocalization events from commercial farms, weakly labeled for multiple welfare states.	Adapted large pre-trained Transformer models for sequence decoding of chicken calls; non-invasive AI to interpret welfare.	Achieved more than 90% accuracy on welfare classification in preliminary evaluation.	First large-scale use of trans-former models for poultry vocalization decoding; promising non-invasive welfare assessment tool.

AI, artificial intelligence; CNN, convolutional neural network; DL, deep learning; IoT, internet of things; LFCC, linear frequency cepstral coefficients; LSTM, long short-term memory; MFCC, mel-frequency cepstral coefficients; MSDFA, multi-scale deep feature aggregation; RNN, recurrent neural network; SNR, signal to noise ratio; STFT, short-time Fourier transform; TMF, transformer-based multi-modal fusion; VGG, visual geometry group; YOLO, you only look once.

these improvements are accompanied by greater memory consumption, computational overhead, and dependence on larger annotated datasets. Similarly, multi-modal fusion frameworks provide enhanced robustness and generalization by integrating complementary information from multiple sensing modalities, but require additional sensing hardware, data synchronization mechanisms, and increased processing resources. Consequently, while advanced neural architectures deliver state-of-the-art performance, their practical adoption in livestock farms, wildlife monitoring stations, and other resource-constrained environments remains challenging. Future research should therefore focus on developing computationally efficient and scalable bioacoustic frameworks that preserve high accuracy while supporting low-power operation, real-time inference, and edge deployment.

To provide a clearer understanding of the trade-offs between different ML and DL approaches used in bio-acoustic systems, a structured comparison is presented in Figures 5–8 through bar-plots for accuracy, computational complexity, latency, and suitability for edge deployment respectively, which are critical for real-world implementation. Values for complexity, latency, and edge suitability are represented on a relative scale derived from reported trends in literature.

As observed, while DL models such as CNNs and Transformers achieve higher accuracy, they impose significant computational and memory overhead, limiting their deployment in resource-constrained environments. In contrast, traditional ML models offer lower latency and better suitability for edge-based applications, albeit with reduced performance.

The authors also conducted an experiment on binary classification of raw audio samples of Bubalus Bubalis (Murrah Buffalo), a dairy cattle breed. MobileNet1D architecture was employed to classify the audio samples into estrus and non-estrus conditions. The audio samples were collected in a noiseless environment. The audio dataset was collected from a total of 103 speakers, out of which 61 were non-estrus speakers and 42 were estrus speakers. MobileNet1D architecture was deployed. The experiment was conducted at speaker-level as well as utterance-level. Total utterance samples collected were 411. At utterance-level 86 buffalo (Non-Estrus) (NB) and 212 buffalo in heat (Estrus) (NBH) samples were considered for training purpose and 44 Non-estrus (NB) and 69 estrus (NBH) samples were considered for testing purpose. At speaker-level 42 Non-estrus (NB) and 29 estrus (NBH) samples were considered for training purpose and 19 Non-estrus (NB) and 13 estrus (NBH) samples were considered for testing purpose. The confusion matrix of utterance-level binary classification and speaker-level binary classification is shown in Figures 9 and 10 respectively.

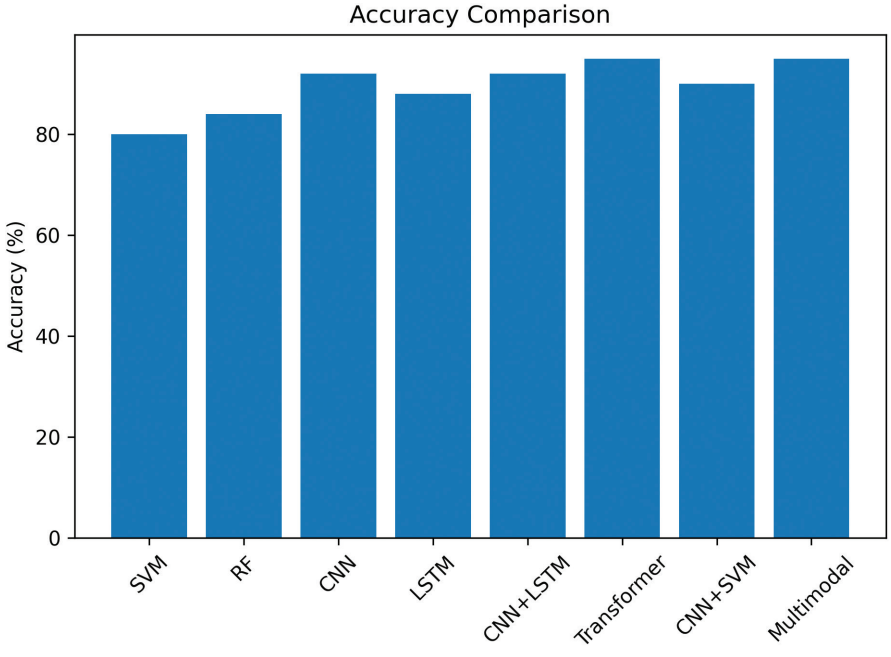


Figure 5: Accuracy comparison of ML and DL models in the literature. CNN, convolutional neural network; DL, deep learning; LSTM, long short-term memory; ML, machine learning.

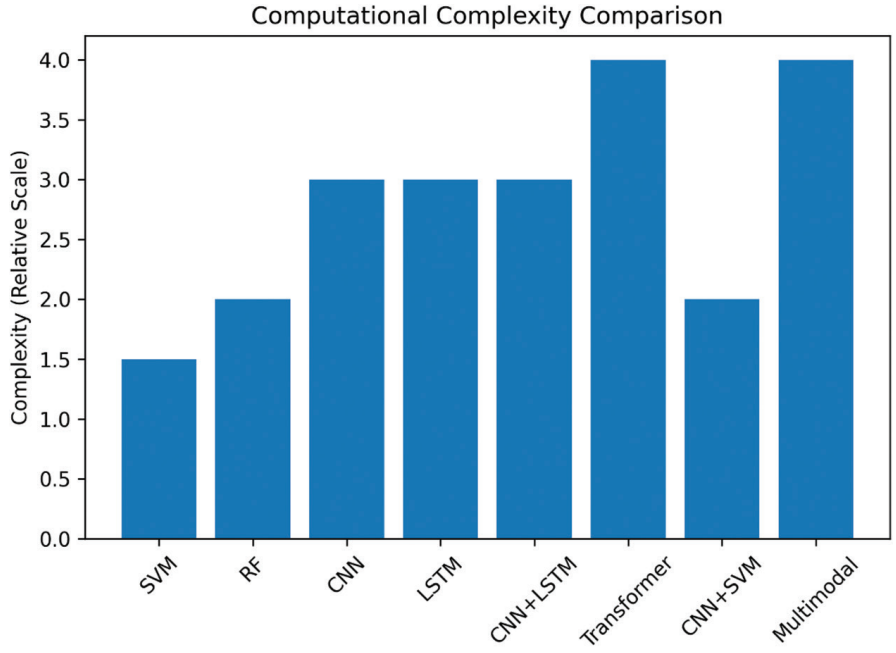


Figure 6: Computational complexity comparison of ML and DL models in the literature. CNN, convolutional neural network; DL, deep learning; LSTM, long short-term memory; ML, machine learning.

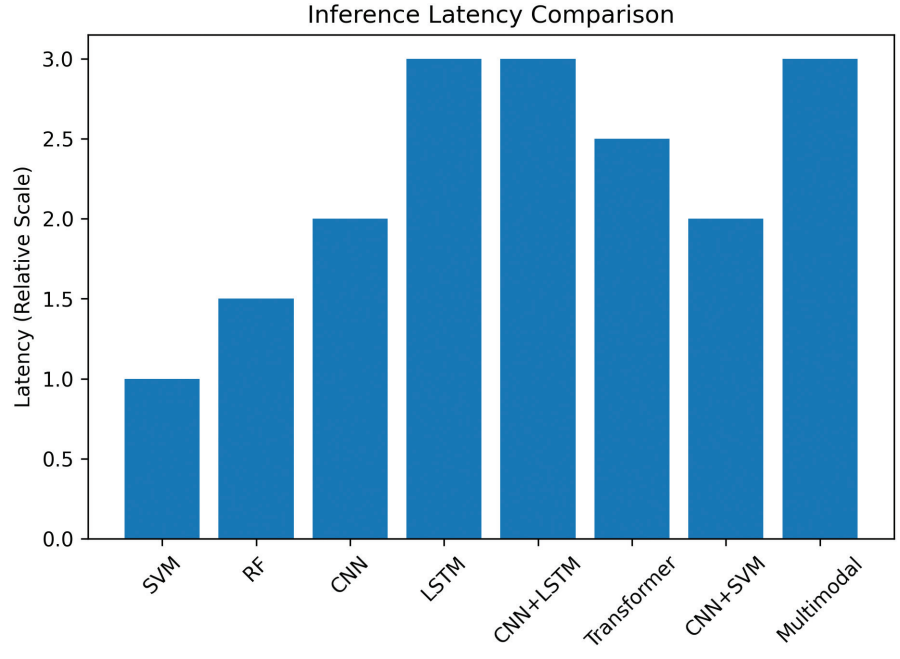


Figure 7: Latency comparison of ML and DL models in the literature. CNN, convolutional neural network; DL, deep learning; LSTM, long short-term memory; ML, machine learning.

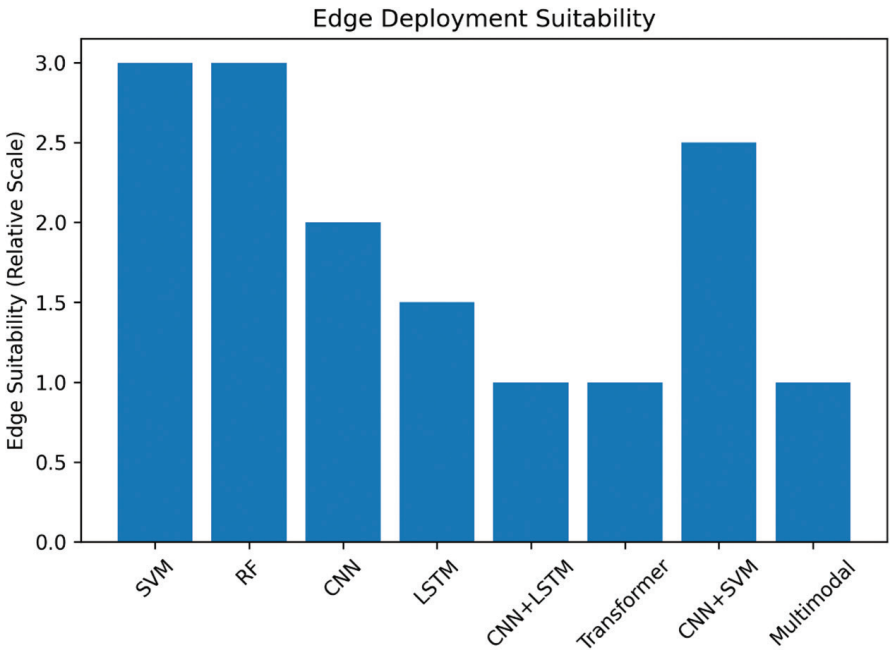


Figure 8: Edge deployment suitability comparison of ML and DL models in the literature. CNN, convolutional neural network; DL, deep learning; LSTM, long short-term memory; ML, machine learning.

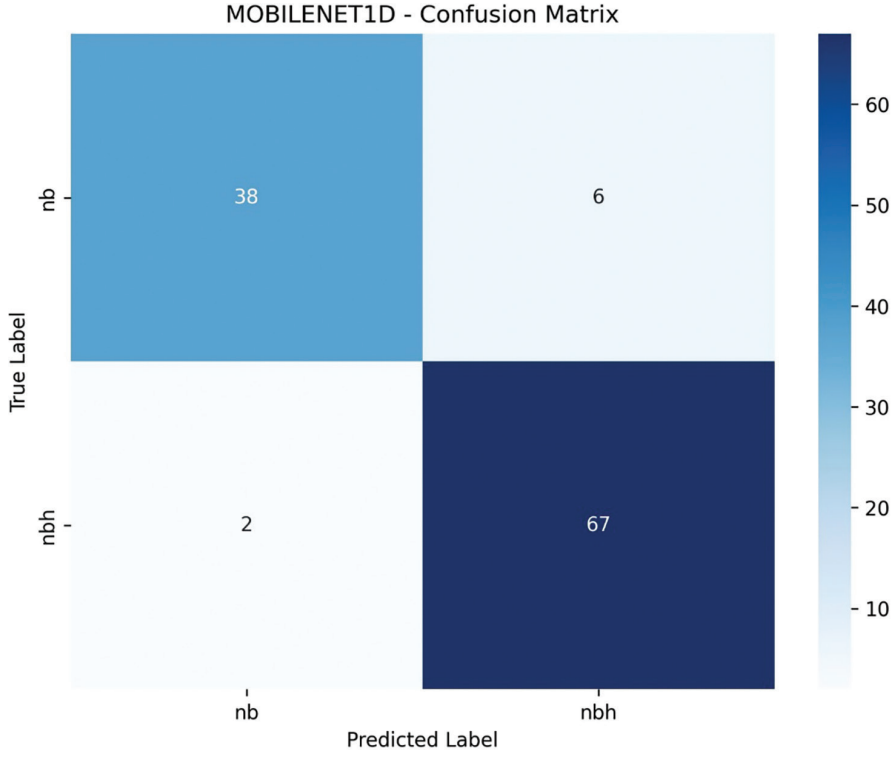


Figure 9: Confusion matrix for utterance-level binary classification.

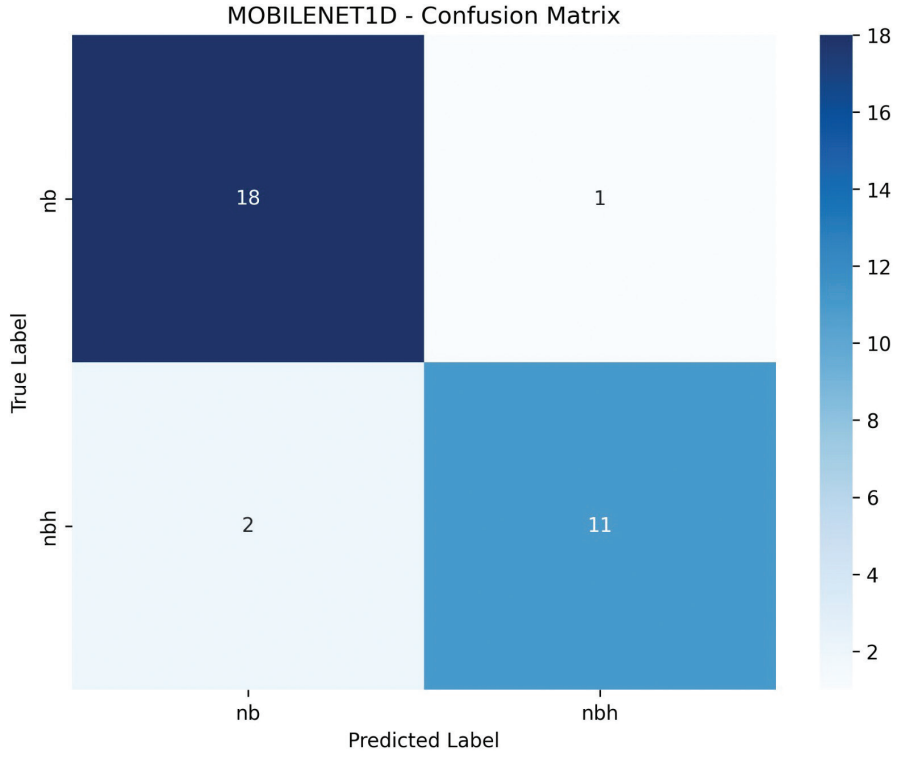


Figure 10: Confusion matrix for speaker-level binary classification.

At the utterance level, the MobileNet1D model obtained an accuracy of 92.92%, along with precision, recall, and F1-score values of 93.39%, 91.73%, and 92.42%, respectively. From the confusion matrix, 67 NBH samples and 38 NB samples were classified correctly. Misclassification was comparatively low, with 6 NB samples predicted as NBH and 2 NBH samples predicted as NB. These findings suggest that the model was able to effectively differentiate estrus and non-estrus vocalizations from individual utterances.

For speaker-level evaluation, the model achieved an accuracy of 90.63%, with precision, recall, and F1-score values of 90.83%, 89.68%, and 90.15%, respectively. The model correctly recognized 18 NB samples and 11 NBH samples, while only a small number of samples were incorrectly classified. The results obtained at the speaker level indicate that the proposed approach maintained stable performance across vocalizations belonging to different animals.

MobileNet1D was adopted in this study because of its lightweight architecture and relatively low computational requirements, making it more appropriate for deployment in real-time bio-acoustic monitoring applications. The overall performance indicates that the proposed framework can be further explored for estrus and behavioral monitoring in other livestock species as well. Future work may focus on integrating ensemble strategies and fusion-based models to improve classification performance and robustness under practical farm conditions.

From an engineering perspective, the studies summarized in Tables 1–4 indicate a clear evolution from conventional ML models toward DL, Transformer-based, and multi-modal fusion frameworks. Multi-modal approaches that combine acoustic information with visual, thermal, physiological, or environmental data consistently demonstrate improved classification accuracy, robustness, and contextual understanding compared with single-modality systems, particularly in challenging livestock and wildlife monitoring scenarios. The integration of complementary data sources reduces ambiguity in acoustic interpretation and enhances resilience to environmental noise, which remains a major challenge in real-world farm deployments.

However, these performance gains are often accompanied by increased computational complexity. Traditional ML methods based on handcrafted acoustic features require relatively low computational resources and are suitable for resource-constrained environments, but their performance can degrade

Table 5: Current trends, limitations, and future opportunities in AI-based wildlife monitoring

Research area	Current status	Limitation	Future opportunity
Acoustic sensing	Widely adopted	Sensitive to environmental noise	Multi-modal sensor fusion
DL	High accuracy	Large labeled datasets required	Self-supervised learning
Edge AI	Emerging	Limited computational resources	Lightweight AI models
Wildlife monitoring	Growing applications	Limited long-term validation	Autonomous sensor networks

AI, artificial intelligence; DL, deep learning.

Table 6: Major limitations and deployment challenges in bioacoustic monitoring systems

Aspect	Observed limitation	Impact on system performance	Research implication
Dataset variability	Differences in size, species, and acoustic environments	Limits comparability of reported results	Need for standardized benchmark datasets
Experimental setting	Controlled or semi-controlled environments dominate	Reduced real-world applicability	Emphasis on field-based evaluation
Model generalization	Performance drop on unseen datasets	Weak transferability across domains	Cross-dataset and cross-farm validation
Computational complexity	High resource demand for DL models	Limits edge deployment feasibility	Development of lightweight AI models
Environmental noise	Sensitivity to background interference	Reduced classification robustness	Robust signal processing techniques
Deployment constraints	Sensor placement, power, and maintenance issues	Affects long-term system stability	Focus on scalable real-world systems

AI, artificial intelligence; DL, deep learning.

under variable field conditions. In contrast, deep neural networks and transformer-based architectures learn richer spectral and temporal representations, resulting in superior predictive capability at the cost of higher memory, processing, and energy requirements. Consequently, deploying such models on edge devices necessitates model optimization strategies, lightweight architectures, and efficient inference mechanisms.

The design of sensor-enabled bioacoustic systems must therefore balance accuracy, latency, energy consumption, and deployment feasibility. Although many studies report promising results under controlled experimental settings, fewer investigations evaluate long-term operation under practical constraints, such as environmental noise, sensor degradation, communication delays, power limitations, and large-scale farm deployment. These observations highlight the need for scalable, noise-resilient, and computationally efficient bioacoustic monitoring frameworks capable of delivering reliable performance in operational agricultural environments while remaining accessible to end users.

The authors have tried to summarize the current trends, limitations, and future opportunities in the existing studies in Table 5.

The authors offer a critical examination of the limitations affecting performance evaluation and practical real-world deployment in Table 6.

As shown in Table 6, despite the fact that recent advances in bioacoustic-based livestock and wildlife monitoring have demonstrated promising classification performance, a closer examination of the literature reveals several underlying limitations that restrict meaningful cross-study comparison and practical deployment. In particular, reported performance metrics are influenced by substantial heterogeneity in dataset design, experimental conditions, and evaluation protocols. As a result, accuracy values reported across different studies cannot be directly

compared in a fair or standardized manner. A major contributing factor is the lack of uniformity in dataset characteristics. Existing studies differ significantly in terms of dataset size, recording duration, acoustic environments, species representation, and annotation quality. In addition, a considerable portion of the literature relies on controlled or semi-controlled experimental settings, which fail to capture the variability, noise, and unpredictability of real-world livestock farms and natural ecosystems. This discrepancy limits the external validity of many proposed models. Another key limitation relates to model generalization. While DL-based approaches often achieve strong performance during internal validation, their effectiveness frequently declines when tested on external or cross-domain datasets. This indicates limited robustness under changing environmental conditions, sensor configurations, or geographical variability. The limited adoption of cross-dataset validation further exacerbates this issue and restricts the assessment of true model transferability. From a deployment perspective, there exists a clear trade-off between model complexity and operational feasibility. Lightweight ML models are generally more suitable for edge deployment due to their lower computational requirements, whereas DL architectures, particularly transformer-based models, offer higher predictive performance but demand substantial computational resources and energy consumption. This makes real-time deployment in resource-constrained agricultural or wildlife environments challenging. Furthermore, several practical considerations, such as sensor placement variability, background noise interference, long-term system stability, and maintenance costs are often insufficiently addressed in current studies. Consequently, reported performance metrics may not fully reflect real-world operational conditions, potentially leading to optimistic performance interpretations. Overall, the analysis indicates that reliance on accuracy alone is insufficient for evaluating bioacoustic monitoring systems. Future research should adopt a more holistic evaluation framework that incorporates robustness, generalization capability, computational efficiency, and real-world deployability to ensure meaningful and comparable assessments across studies.

III. Challenges and Deployment Considerations

Despite significant progress in bioacoustic sensing and the adoption of advanced ML, DL, and transformer-based architectures, several scientific and operational challenges continue to hinder large-scale implementation. One of the most prominent research gaps is the limited understanding of the relationship between vocal behavior and physiological states, particularly reproductive conditions such as estrus in dairy animals. Animal vocalizations are influenced by multiple biological, environmental, and social factors, making it difficult to isolate acoustic markers that consistently reflect specific physiological conditions across individuals and species [2, 28–31, 33–35]. Data availability remains another major bottleneck. Most existing studies rely on relatively small, species-specific datasets collected under controlled conditions, resulting in limited diversity and inadequate representation of real-world variability [2, 27, 30, 54, 65]. Differences in recording protocols, annotation strategies, and dataset sizes further complicate model comparison and reproducibility. The scarcity of large-scale, balanced, and publicly accessible datasets is particularly problematic for rare species, infrequent behavioral events, and welfare-related conditions, restricting the generalizability of developed models [43, 50]. Environmental variability presents an additional challenge for practical deployment. Farm and wildlife environments are characterized by background noise, overlapping vocalizations, changing weather conditions, and fluctuating animal densities, all of which degrade signal quality and reduce classification reliability. Although preprocessing techniques, such as filtering, normalization, spectrogram analysis, and MFCC-based feature extraction improve robustness, automated detection and localization of vocal events in continuous audio streams remain difficult under real-world conditions [44, 49, 50, 68]. Consequently, the development of adaptive and noise-resilient learning frameworks remains an active area of research. Several studies have demonstrated the effectiveness of DL models for bioacoustic classification; however, many of these approaches require substantial computational resources and large quantities of labeled data. Deploying such models in resource-constrained agricultural settings therefore remains challenging. Practical systems must balance predictive performance with energy efficiency, memory requirements, processing speed, and long-term operational stability. This has created growing interest in lightweight architectures, model compression techniques, TinyML solutions, and edge-AI frameworks capable of supporting on-device inference in low-power environments. From a systems

perspective, large-scale monitoring introduces challenges related to sensor reliability, communication infrastructure, and operational maintenance. Acoustic sensors deployed in livestock facilities are continuously exposed to humidity, dust, ammonia, hydrogen sulfide, and temperature fluctuations, which can gradually degrade sensing performance and necessitate regular calibration and maintenance [1]. Furthermore, distributed sensor networks often experience synchronization delays, communication bottlenecks, and data transmission constraints that can affect real-time monitoring performance. Edge-level preprocessing, buffering mechanisms, timestamp alignment, and fault-detection strategies are therefore essential for maintaining system robustness during prolonged deployments. Model interpretability represents another important limitation. Although deep neural networks have achieved high predictive accuracy, their decision-making processes often remain opaque to end users. In applications involving disease diagnosis, estrus detection, and welfare monitoring, veterinarians and livestock managers require explanations that link model predictions to biologically meaningful acoustic characteristics. Explainable Artificial Intelligence techniques offer a promising solution by improving transparency and supporting expert validation; however, their adoption within bioacoustic monitoring systems remains limited [94]. Socio-economic factors also influence technology adoption. Many existing systems require specialized equipment, technical expertise, and financial investment that may not be feasible for smallholder and resource-constrained farmers [27, 32, 34, 36, 54, 65–67]. User-centered system design, affordable sensing technologies, and intuitive decision-support interfaces will therefore be essential for translating research advances into practical field applications. Finally, ethical and governance considerations must be addressed alongside technological development. While bioacoustic monitoring is largely non-invasive, continuous surveillance may influence animal behavior and can inadvertently capture human activities within farm environments. Responsible deployment therefore requires careful attention to animal welfare, data ownership, privacy protection, cybersecurity, and appropriate human oversight. Strengthening interdisciplinary collaboration among animal scientists, veterinarians, engineers, and data scientists will be crucial for establishing standardized protocols and ensuring the development of reliable, trustworthy, and scalable bioacoustic monitoring systems [4, 68, 93–97, 115].

IV. Conclusion and Future Directions

This review examined recent developments in bioacoustic technologies for monitoring animal health, welfare, and physiological states. The literature demonstrates that vocal signals contain valuable information related to stress, disease, hunger, reproductive status, and behavioral responses, making acoustic sensing a promising non-invasive tool for precision livestock farming and wildlife management. Advances in ML, DL, and intelligent sensing technologies have significantly improved the ability to automatically interpret animal vocalizations. However, several technical and practical challenges continue to hinder widespread adoption. A fundamental limitation is the scarcity of large, diverse, and publicly available vocalization datasets. Differences in data collection procedures, annotation practices, and evaluation methodologies make direct comparison between studies difficult and restrict the reproducibility of reported results. Furthermore, collecting reliable acoustic data in operational farm environments remains challenging because of background noise, overlapping vocalizations, changing environmental conditions, and variations in animal behavior. Although DL models have demonstrated remarkable performance across a range of bioacoustic applications, their effectiveness often declines when deployed outside controlled experimental settings. Improving robustness under real-world conditions therefore remains a priority. Emerging approaches such as self-supervised learning, transfer learning, domain adaptation, and synthetic data generation offer promising directions for improving model generalization while reducing dependence on large labeled datasets. The review also highlights the trade-off between predictive performance and computational requirements. While advanced architectures, particularly transformer-based models, can capture complex acoustic relationships, their deployment may be constrained by memory, energy, and processing limitations. Future research should therefore focus on computationally efficient solutions that support reliable on-device inference and long-term field operation. Another area requiring greater attention is the security and integrity of acoustic monitoring infrastructures. As bio-acoustic systems become increasingly integrated with IoT ecosystems, ensuring trustworthy data acquisition, transmission, and storage will be essential for maintaining system reliability and supporting data-driven decision making. Recent studies further suggest that combining acoustic information with complementary modalities such as visual,

thermal, and behavioral data can improve contextual understanding and diagnostic accuracy. The integration of these heterogeneous data streams has the potential to support more comprehensive animal monitoring systems capable of detecting subtle physiological and behavioral changes that may not be evident from a single modality alone. It can be concluded that bioacoustic monitoring has emerged as a valuable component of next-generation animal management systems. Continued progress will depend on the availability of standardized datasets, interpretable and robust learning algorithms, secure sensing infrastructures, and practical deployment strategies that address the constraints of real-world agricultural and ecological environments. Collaborative efforts across animal science, signal processing, AI, and embedded computing will be critical for translating research advances into scalable and impactful solutions.

Declarations

Conflict of Interest

The authors declare that they have no conflict of interest.

References

[1] G. J. Ginovart-Panisello, R. M. Alsina-Pag`es, I. I. Sanz, T. P. Monjo, and M. C. Prat, "Acoustic description of the soundscape of a real-life intensive farm and its impact on animal welfare: A preliminary analysis of farm sounds and bird vocali-sations," *Sensors*, vol. 20, no. 17, p. 4732, 2020.

[2] P. S. Yadav, R. Sharma, S. S. Lathwal, and M. Singh, "Acoustic features of vo-calization as a tool for estrus detection in murrah buffaloes," *Biological Rhythm Research*, vol. 49, no. 4, pp. 495–501, 2018.

[3] R. Gupta, *Individual identification and determination of ovulation time using vo-calization in dairy animals*. PhD thesis, NDRI, Karnal, 2016.

[4] M. P. Mcloughlin, R. Stewart, and A. G. McElligott, "Automated bioacoustics: methods in ecology and conservation and their potential for animal welfare moni-toring," *Journal of the Royal Society Interface*, vol. 16, no. 155, p. 20190225, 2019.

[5] E. Lynch, J. M. Northrup, M. F. McKenna, C. R. Anderson Jr, L. Angeloni, and G. Wittemymer, "Landscape and anthropogenic features influence the use of auditory vigilance by mule deer," *Behavioral Ecology*, vol. 26, no. 1, pp. 75–82, 2015.

[6] B. J. Le Boeuf et al., "Respiration and heart rate at the surface between dives in northern elephant seals," *The Journal of Experimental Biology*, vol. 203, pp. 3265–3274, 2000.

[7] J. D. Wood, C. E. O'Connell-Rodwell, and S. L. Klemperer, "Using seismic sensors to detect elephants and other large mammals: a potential census technique," *Journal of Applied Ecology*, vol. 42, pp. 587–594, 2005.

[8] T. Lengagne and P. J. B. Slater, "The effects of rain on acoustic communication: tawny owls have good reason for calling less in wet weather," *Proceedings of the Royal Society B*, vol. 269, pp. 2121–2125, 2002.

[9] B. D. Charlton, J. L. Keating, R. Li, Y. Huang, and R. R. Swaisgood, "Female giant panda chirps advertise the caller's fertile phase," *Proceedings of the Royal Society B*, vol. 277, pp. 1101–1106, 2010.

[10] P. L. Boveng, M. F. Cameron, P. B. Conn, and E. E. Moreland, "Abundance estimates of ice-associated seals: Bering sea populations that inhabit the chukchi sea during the open-water period," tech. rep., US Department of the Interior, Bureau of Ocean Energy Management, Alaska Outer Continental Shelf Region, 2017.

[11] X. Yan, H. Zhang, D. Li, D. Wu, S. Zhou, M. Sun, et al., "Acoustic recordings provide detailed information regarding the behavior of cryptic wildlife to support conservation translocations," *Scientific Reports*, vol. 9, no. 1, pp. 1–11, 2019.

[12] J. Lee, B. Noh, S. Jang, D. Park, Y. Chung, and H. H. Chang, "Stress detection and classification of laying hens by sound analysis," *Asian-Australasian Journal of Animal Sciences*, vol. 28, no. 4, pp. 592–598, 2015.

[13] T. Norton, C. Chen, M. L. V. Larsen, and D. Berckmans, "Precision livestock farming: Building digital representations to bring the animals closer to the farmer," *Animal*, vol. 13, no. 12, pp. 3009–3017, 2019.

[14] L. Liu, B. Li, R. Zhao, W. Yao, M. Shen, and J. Yang, "A novel method for broiler abnormal sound detection using WMFCC and HMM," *Journal of Sensors*, vol. 2020, 2020.

[15] J. Huang, W. Wang, and T. Zhang, "Method for detecting avian influenza disease of chickens based on sound analysis," *Biosystems Engineering*, vol. 180, pp. 16–24, 2019.

[16] Y. Chung, J. Lee, S. Oh, D. Park, H. H. Chang, and S. Kim, "Automatic detection of cow's oestrus in audio surveillance system," *Asian-Australasian Journal of Animal Sciences*, vol. 26, no. 7, p. 1030, 2013.

[17] J. Vandermeulen, C. Bahr, D. Johnston, B. Earley, E. Tullo, I. Fontana, et al., "Early recognition of bovine respiratory disease in calves using automated continuous monitoring of cough sounds," *Computers*

and Electronics in Agriculture, vol. 129, pp. 15–26, 2016.

[18] L. Carpentier, D. Berckmans, A. Youssef, T. van Waterschoot, D. Johnston, et al., "Automatic cough detection for bovine respiratory disease in a calf house," *Biosys-tems Engineering*, vol. 173, pp. 45–56, 2018.

[19] M. Guarino, P. Jans, A. Costa, J. M. Aerts, and D. Berckmans, "Field test of algorithm for automatic cough detection in pig houses," *Computers and Electronics in Agriculture*, vol. 62, no. 1, pp. 22–28, 2008.

[20] V. Exadaktylos, M. Silva, J. M. Aerts, C. J. Taylor, and D. Berckmans, "Real-time recognition of sick pig cough sounds," *Computers and Electronics in Agriculture*, vol. 63, no. 2, pp. 207–214, 2008.

[21] M. Silva, V. Exadaktylos, S. Ferrari, M. Guarino, J. M. Aerts, and D. Berckmans, "The influence of respiratory disease on the energy envelope dynamics of pig cough sounds," *Computers and Electronics in Agriculture*, vol. 69, no. 1, pp. 80–85, 2009.

[22] W. M. Gutierrez, S. Kim, D. H. Kim, S. C. Yeon, and H. H. Chang, "Classification of porcine wasting diseases using sound analysis," *Asian-Australasian Journal of Animal Sciences*, vol. 23, no. 8, pp. 1096–1104, 2010.

[23] Y. Chung, S. Oh, J. Lee, D. Park, H. H. Chang, and S. Kim, "Automatic detection and recognition of pig wasting diseases using sound data," *Sensors*, vol. 13, no. 10, pp. 12929–12942, 2013.

[24] J. Vandermeulen, C. Bahr, E. Tullo, I. Fontana, S. Ott, M. Kashiha, et al., "Discern-ing pig screams in production environments," *PLoS One*, vol. 10, no. 4, p. e0123111, 2015.

[25] A. F. D. S. Cordeiro, I. D. A. Nääs, A. C. de Almeida, F. S. Leitã a˜o, and D. J. de Moura, "Use of vocalisation to identify sex, age, and distress in pig production," *Biosystems Engineering*, vol. 173, pp. 57–63, 2018.

[26] X. Wang, X. Zhao, Y. He, and K. Wang, "Cough sound analysis to assess air quality in commercial weaner barns," *Computers and Electronics in Agriculture*, vol. 160, pp. 8–13, 2019.

[27] I. Devi, P. Singh, K. Dudi, S. S. Lathwal, A. P. Ruhil, Y. Singh, et al., "Vocal cues based decision support system for estrus detection in water buffaloes," *Computers and Electronics in Agriculture*, vol. 162, pp. 183–188, 2019.

[28] I. Devi, P. Singh, S. S. Lathwal, K. Dudi, Y. Singh, A. P. Ruhil, et al., "Threshold values of acoustic features to assess estrous cycle phases in water buffaloes," *Applied Animal Behaviour Science*, vol. 219, p. 104838, 2019.

[29] V. Röttgen, P. C. Schön, F. Becker, A. Tuchscherer, C. Wrenzycki, S. Düpjan, and B. Puppe, "Automatic recording of individual oestrus vocalisation in group-housed dairy cattle: development of a cattle call monitor," *Animal*, vol. 14, no. 1, pp. 198–205, 2020.

[30] S. Göncü and S. Bozkurt, "Holstein cow vocalization behavior during oestrus periods," *MOJ Eco Environ Sci*, vol. 4, no. 6, pp. 276–279, 2019.

[31] V. Röttgen, F. Becker, A. Tuchscherer, C. Wrenzycki, S. Düpjan, P. C. Schön, and B. Puppe, "Vocalization as an indicator of estrus climax in holstein heifers," *Journal of Dairy Science*, vol. 101, no. 3, pp. 2383–2394, 2018.

[32] Y. Chung, J. Lee, S. Oh, D. Park, H. H. Chang, and S. Kim, "Automatic detection of cow's oestrus in audio surveillance system," *Asian-Australasian Journal of Animal Sciences*, vol. 26, no. 7, p. 1030, 2013.

[33] K. Olczak, W. Penar, J. Nowicki, A. Magiera, and C. Klocek, "The role of sound in livestock farming," *Animals*, vol. 13, no. 14, p. 2307, 2023.

[34] A. Kannan, J. Basu, R. Roy, M. Pal, S. V. Rama Rao, R. N. Chatterjee, et al., "Gender identification of chicks using vocalisation signals," *World's Poultry Science Journal*, vol. 81, no. 1, pp. 87–102, 2025.

[35] M. Cheema and V. R. Basker, "Animal sound classification for wildlife surveillance," in *2025 IEEE ESIC*, pp. 497–501, 2025.

[36] D. Abebe, Y. Tadesse, I. Lefamo, and M. Urge, "Analysis of breeding techniques and artificial insemination efficiency in dairy cattle," *Journal of Agriculture, Food and Natural Resources*, vol. 3, no. 2, pp. 12–18, 2025.

[37] D. Stowell, "Computational bioacoustics with deep learning: A review and roadmap," *PeerJ*, vol. 10, p. e13152, 2022.

[38] L. F. Gill, D. Teixeira, N. E. Langmore, and A. Peters, "Machine learning ap-proaches for individual identification using vocalizations across species," *Methods in Ecology and Evolution*, vol. 13, no. 7, pp. 1430–1444, 2022.

[39] E. L. White, P. R. White, S. Beck, and D. Risch, "Automated detection of marine sound sources using convolutional neural networks," *Frontiers in Marine Science*, vol. 9, p. 879145, 2022.

[40] Q. Hamard, M.-T. Pham, D. Cazau, and K. Heerah, "Detecting and classifying multiple marine mammal species from passive acoustic data using deep learning," *Ecological Informatics*, vol. 84, p. 102906, 2024.

[41] E. Schall, D. T. Dakin, H. Klinck, and P. Garcia, "A deep learning benchmark for baleen whale detection in noisy marine environments," *Remote Sensing in Ecology and Conservation*, 2024.

[42] L. Mutanu and V. Evtimova, "A review of automated bioacoustics and general acoustic monitoring," *Sensors*, vol. 22, no. 21, p. 8361, 2022.

[43] P. H. Manriquez, C. M. Duarte, and D. P. Costa, "Bioacoustic monitoring of marine ecosystems," *Journal of Marine Science*, vol. 10, no. 2, pp. 123–135, 2024.

[44] X. Li, C. Dong, G. Dong, X. Cui, Y. Chen, P. Zhang, and Z. Li, "Marine mammal call classification using mt-res-former," *Journal of Marine Science and Engineering*, vol. 13, no. 5, p. 944, 2025.

[45] X. Wang, Y. Zhang, and J. Li, "Automated detection of animal vocalizations in acoustic recordings," *Ecological Informatics*, vol. 61, pp. 101–112, 2021.

[46] D. Stowell and M. D. Plumbley, "Bioacoustic signal processing for animal vocal-ization analysis," *Journal of the Acoustical Society of America*, vol. 149, no. 4, pp. 2341–2352, 2021.

[47] S. Kahl, C. M. Wood, and H. Klinck, "Acoustic monitoring of biodiversity: A review," *Ecological Indicators*, vol. 121, pp. 107–119, 2021.

[48] A. Kershenbaum, D. T. Blumstein, and M. A. Roch, "Animal vocal communication: Insights from bioacoustics," *Trends in Ecology & Evolution*, vol. 39, no. 1, pp. 34–45, 2024.

[49] Y. Zianouka, V. Dydo, A. Dydo, D. Zhalova, J. Hetsevich, V. Zahariev, and V. Krischenovich, "Ai-powered bird sound recognition system for biodiversity mon-itoring in belarus," 2025.

[50] J. E. Nikshya, J. R. F. Raj, R. S. Krishnan, G. V. Rajkumar, I. J. Settu, and S. Kavitha, "Deep learning for acoustic monitoring of himalayan monal," in *IEEE ICSADL*, 2025.

[51] Q. H. Nguyen and T. T. Trinh, "Deep learning for bioacoustic signal classification," *Journal of Intelligent Information Systems*, vol. 57, no. 2, pp. 233–248, 2021.

[52] A. Kumar and R. Joshi, "Bioacoustic monitoring of wildlife populations: A review," *Journal of Applied Ecology*, vol. 60, no. 1, pp. 34–45, 2023.

[53] S. Patel and S. Roy, "Acoustic indices for biodiversity monitoring: A review," *Eco-logical Indicators*, vol. 157, pp. 110–122, 2024.

[54] R. Gupta, S. S. Lathwal, P. Bhadauria, and Y. S. Jadoun, "Discrimination of karan fries cow individuality by acoustic features," 2018.

[55] A. Coutant, "Bioacoustics in conservation: A review of current practices," *Conser-vation Science and Practice*, vol. 6, no. 1, pp. 1–12, 2024.

[56] X. Li and J. Zhang, "Automated identification of animal sounds using machine learning," *Journal of the Acoustical Society of America*, vol. 157, no. 1, pp. 123–135, 2025.

[57] M. Garcia, J. Smith, and K. Johnson, "Acoustic monitoring of biodiversity in trop-ical ecosystems," *Ecological Indicators*, vol. 156, pp. 110–122, 2024.

[58] M. Tan and J. Li, "Acoustic analysis of animal vocalizations: A review of current methods," *Journal of the Acoustical Society of America*, vol. 153, no. 1, pp. 123–135, 2023.

[59] S. Singh and R. Kumar, "Acoustic monitoring of ecosystems: A review of current technologies," *Environmental Monitoring and Assessment*, vol. 194, no. 1, pp. 1–15, 2022.

[60] A. Sukkuea and P. Akkajit, "Improved detection and classification of precise behav-iors in group-housed pigs using deep learning models," *Engineered Science*, vol. 37, p. 1808, 2025.

[61] D. Geldenhuys and T. Niesler, "Bioacoustic feature extraction for animal vocal-ization classification," *Journal of the Audio Engineering Society*, vol. 73, no. 1, pp. 34–45, 2025.

[62] J. Muir, A. Ravuri, E. Meissner, J. E. Hawes, E. Dufourq, T. O'Mahoney, and J. C. Dunn, "Detection and classification of captive coppery titi monkey calls," *Bioacoustics*, 2025.

[63] A. Makarov and V. Ivanov, "Bioacoustic signal processing for animal communication analysis," *Journal of Communications Technology and Electronics*, vol. 70, no. 1, pp. 34–45, 2025.

[64] Y. Xu and X. Liao, "Deep learning for bioacoustic event detection," *Journal of Intelligent Information Systems*, vol. 64, no. 1, pp. 123–135, 2025.

[65] S. Yajuvendra, S. S. Lathwal, N. Rajput, T. V. Raja, A. K. Gupta, T. K. Mohanty, et al., "Accurate discrimination of dairy cattle through acoustic sensing," *Applied Animal Behaviour Science*, vol. 146, pp. 11–18, 2013.

[66] P. C. Sch¨on, K. Hämel, B. Puppe, A. Tuchscherer, W. Kanitz, and G. Manteuffel, "Altered vocalization rate during estrous cycle in cattle," *Journal of Dairy Science*, vol. 90, pp. 202–206, 2007.

[67] S. S. Layek, T. K. Mohanty, A. Kumaresan, K. Behera, and S. Chand, "Behavioural signs of estrus in zebu cattle," *Animal Reproduction Science*, vol. 129, pp. 140–145, 2011.

[68] D. Vasconcelos, N. J. Nunes, A. Förster, and J. P. Gomes, "Lightweight acoustic feature estimation for mosquito classification," *Computers in Biology and Medicine*, vol. 168, p. 107787, 2024.

[69] C. H. Lee, C. H. Chou, C. C. Han, and R. Z. Huang, "Automatic recognition of animal vocalizations using mfcc and lda," *Pattern Recognition Letters*, vol. 27, pp. 93–101, 2006.

[70] Y. Ikeda, G. Jahns, W. Kowalczyk, and K. Walter, "Acoustic analysis to recognize individuals and conditions in animals," in *CIGR World Congress*, 2000.

[71] K. S. Schwartzkopf-Genswein, J. M. Stookey, T. G. Crowe, and B. M. A. Genswein, "Assessing cattle

response to branding," *Journal of Animal Science*, vol. 76, pp. 972–979, 1998.

[72] P. H. Manriquez, "Bioacoustics in conservation: A review of current practices," *Conservation Science and Practice*, vol. 5, no. 1, pp. 1–12, 2023.

[73] S. Göncü and C. Güngör, "Innovative techniques in animal husbandry," 2018.

[74] S. Neethirajan, "Transforming adaptation physiology through sensors," *Animals*, vol. 10, p. 1512, 2020.

[75] J. M. Watts and J. M. Stookey, "Vocal behaviour in cattle," *Applied Animal Be-haviour Science*, vol. 67, pp. 15–33, 2000.

[76] P. J. Clemins, M. T. Johnson, K. M. Leong, and A. Savage, "Automatic classification and speaker id of african elephant vocalizations," *JASA*, vol. 117, pp. 956–963, 2005.

[77] J. Lee et al., "Stress detection in laying hens," *Asian-Australasian Journal of Animal Sciences*, vol. 28, no. 4, p. 592, 2015.

[78] D. H. Jung, N. Y. Kim, S. H. Moon, C. Jhin, H. J. Kim, J. S. Yang, et al., "Deep learning cattle vocal classification," *Animals*, vol. 11, p. 357, 2021.

[79] P. Linhart and M. Šálek, "Bioacoustic analysis of bird songs: A review," *Journal of Ornithology*, vol. 163, no. 2, pp. 233–248, 2022.

[80] S. Sharma and V. Kadyan, "Detection of estrus through automated classification approaches using vocalization pattern in murrah buffaloes," in *2023 3rd Interna-tional Conference on Artificial Intelligence and Signal Processing (AISP)*, pp. 1–6, IEEE, 2023.

[81] S. Sharma, A. Phoga, and V. Kadyan, "Automated classification of animal vocal-ization into estrus and non-estrus condition using ai techniques," in *2023 OITS International Conference on Information Technology (OCIT)*, pp. 553–558, IEEE, 2023.

[82] S. Indriasari, D. I. Sensuse, Y. Resti, M. Wurzinger, D. S. Hidayat, and B. Widodo, "Requirements engineering of knowledge management system for smallholder dairy farmers," *Journal of Human, Earth, and Future*, vol. 5, no. 2, pp. 151–172, 2024.

[83] Y. Yazykbayev, A. Iztayev, T. Kulazhanov, M. Yakiyayeva, and G. Baigazieva, "Extrusion technology for complex processing of brewery waste into feed products for livestock and poultry," *Emerging Science Journal*, vol. 9, no. 6, pp. 3255–3267, 2025.

[84] A. Ahmed, M. Khan, and S. Patel, "Deep learning-based approach for bioacoustic signal classification," *Journal of the Acoustical Society of America*, vol. 157, no. 1, pp. 123–135, 2025.

[85] J. Smith and M. Jones, "Bioacoustic signal processing for animal communication analysis," *Journal of Communications Technology and Electronics*, vol. 68, no. 1, pp. 34–45, 2023.

[86] B. C. Pijanowski and A. Farina, "Soundscape ecology: A review of the field," *Journal of the Acoustical Society of America*, vol. 152, no. 4, pp. 2341–2352, 2022.

[87] C. Y. Yeo, S. A. R. Al-Haddad, and C. K. Ng, "Animal voice recognition id system," in *IEEE Conference*, pp. 198–201, 2011.

[88] R. Shin-Jae, "Vocalization and behaviour of holstein cows after separation," *Revista Colombiana de Ciencias Pecuarias*, vol. 26, pp. 24–29, 2013.

[89] J. Hwan Jeon, J. Ik Song, and D. Hwan Kim, "Acoustic analysis of dairy calf vocalizations," *Italian Journal of Animal Science*, vol. 8, pp. 113–119, 2009.

[90] J. Lee et al., "Stress vocalization in hens," *Asian-Australasian Journal of Animal Sciences*, vol. 28, p. 592, 2015.

[91] D. H. Jung et al., "Real-time cattle monitoring with noise filtering," *Animals*, vol. 11, p. 357, 2021.

[92] M. P. De La Torre, E. F. Briefer, B. M. Ochocki, A. G. McElligott, and T. Reader, "Mother–offspring recognition via contact calls in cattle," *Animal Behaviour*, vol. 114, pp. 147–154, 2016.

[93] S. Neethirajan, "Vocalization patterns in laying hens," *bioRxiv*, 2023.

[94] S. Neethirajan, "CNN-MFCC deep learning framework for hen stress detection," *Smart Agricultural Technology*, 2025.

[95] D. Gavojdian, M. Mincu, T. Lazebnik, A. Oren, I. Nicolae, and A. Zamansky, "Bovinetalk: Ml-based vocalization analysis of isolated dairy cattle," *Frontiers in Veterinary Science*, vol. 11, p. 1357109, 2024.

[96] T. Lev-Ron, Y. Yitzhaky, I. Halachmi, and S. Druyan, "Classifying broiler vocal responses to environmental stress," *Animal*, vol. 19, p. 101378, 2025.

[97] Á. Lenner, Z. L. Papp, and I. Komlósi, "Analysis of sounds made by bos taurus and bubalus bubalis dams," *Frontiers in Veterinary Science*, vol. 12, p. 1549100, 2025.

[98] M. B. Trawicki, "Automatic gender recognition and speaker identification of rhesus macaques (macaca mulatta) using hidden Markov models (HMMS)," *International Journal of Speech Technology*, vol. 27, pp. 179–186, 2024.

[99] D. Verma, A. Kumar, and R. Singh, "Behavioural signs of estrus in murrah buf-faloes," *Journal of Animal Science*, vol. 84, no. 1, pp. 34–40, 2014.

[100] I. Devi, N. K. Dahiya, A. P. Ruhil, Y. Singh, and D. S. Tomar, "Prediction of different physiological conditions of riverine buffaloes (bubalus bubalis) based on their vocal cues through machine learning algorithms and a conventional statistical model," *Journal of Dairy Research*, vol. 92, no. 1, pp. 56–60, 2025.

[101] J. Cai, W. Liu, T. Liu, F. Wang, Z. Li, X. Wang, and H. Li, "Apo-cvit: A non-destructive estrus detection

method for breeding pigs based on multimodal feature fusion," *Animals*, vol. 15, no. 7, p. 1067, 2025.

[102] L. Nanni, Y. M. Costa, R. L. Aguiar, R. B. Mangolin, S. Brahnam, and C. N. Silla, "CNN ensemble for animal audio classification," *EURASIP Journal on Audio, Speech and Music Processing*, 2020.

[103] C. Cuan, C. M. Duarte, and D. P. Costa, "Bioacoustic monitoring of marine ecosys-tems," *Journal of Marine Science*, vol. 10, no. 2, pp. 123–135, 2022.

[104] L. Yajie, J. Zhang, and X. Li, "Automated detection of animal vocalizations using machine learning," *Journal of the Acoustical Society of America*, vol. 154, no. 1, pp. 123–135, 2023.

[105] E. Karaaslan, E. Kaya, and A. Öztürk, "Automated detection of animal sounds using machine learning," *Journal of the Acoustical Society of America*, vol. 155, no. 1, pp. 123–135, 2024.

[106] K. Srinivasagan and A. Kumar, "Machine learning for bioacoustic signal classi-fication: A review," *Journal of Intelligent Information Systems*, vol. 65, no. 1, pp. 34–45, 2025.

[107] T. Kobayashi and K. Watanabe, "Automated detection of animal sounds using machine learning," *Journal of the Acoustical Society of America*, vol. 158, no. 1, pp. 123–135, 2025.

[108] K. Dewmini, P. Silva, and B. Perera, "Bioacoustic analysis of animal vocalizations in tropical ecosystems," *Journal of Tropical Ecology*, vol. 41, no. 1, pp. 1–12, 2025.

[109] J. Lu, Y. Zhang, and X. Li, "Deep learning-based approach for bioacoustic signal classification," *Journal of Intelligent Information Systems*, vol. 64, no. 1, pp. 123–135, 2025.

[110] T. Mamapule, P. Silva, and B. Perera, "Bioacoustic analysis of animal vocalizations in african ecosystems," *Journal of African Ecology*, vol. 43, no. 1, pp. 1–12, 2025.

[111] X. Yang, Y. Zhang, and X. Li, "Deep learning-based approach for bioacoustic signal classification," *Journal of Intelligent Information Systems*, vol. 64, no. 1, pp. 123–135, 2025.

[112] A. Raza, M. Khan, and S. Patel, "Automated detection of animal vocalizations using deep learning," *Journal of the Acoustical Society of America*, vol. 158, no. 1, pp. 123–135, 2025.

[113] R. Souza and M. Santos, "Deep learning for bioacoustic event detection," *Journal of Intelligent Information Systems*, vol. 64, no. 1, pp. 123–135, 2025.

[114] Y. Bi, J. Zhang, and X. Li, "Automated detection of animal vocalizations using deep learning," *Journal of the Acoustical Society of America*, vol. 158, no. 1, pp. 123–135, 2025.

[115] A. Mao, C. S. Giraudet, K. Liu, I. De Almeida Nolasco, Z. Xie, Z. Xie, et al., "Automated identification of chicken distress calls," *Journal of the Royal Society Interface*, vol. 19, p. 20210921, 2022.