

Review

Not peer-reviewed version

---

# Giving Cows a Digital Voice – AI-Enabled Bioacoustics and Smart Sensing in Precision Livestock Management

---

[Mayuri Kate](#) and [Suresh Neethirajan](#) \*

Posted Date: 23 May 2025

doi: 10.20944/preprints202505.1857.v1

Keywords: bovine bioacoustics; precision livestock farming; sensor fusion; explainable AI; animal welfare outcomes



Preprints.org is a free multidisciplinary platform providing preprint service that is dedicated to making early versions of research outputs permanently available and citable. Preprints posted at Preprints.org appear in Web of Science, Crossref, Google Scholar, Scilit, Europe PMC.

Copyright: This open access article is published under a Creative Commons CC BY 4.0 license, which permit the free download, distribution, and reuse, provided that the author and preprint are cited in any reuse.

Disclaimer/Publisher's Note: The statements, opinions, and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions, or products referred to in the content.

Review

# Giving Cows a Digital Voice—AI-Enabled Bioacoustics and Smart Sensing in Precision Livestock Management

Mayuri Kate <sup>1</sup> and Suresh Neethirajan <sup>1,2,\*</sup>

<sup>1</sup> Faculty of Computer Science, Dalhousie University, 6050 University Ave, Halifax, NS B3H 1W5, Canada

<sup>2</sup> Faculty of Agriculture, Dalhousie University, 62 Cumming Dr, Bible Hill, NS B2N 5E3, Canada

\* Correspondence: Sureshrajada@dal.ca

**Abstract:** Cattle express their physiological and emotional states through vocalizations, often long before visible behavioral symptoms emerge. This review critically examines the evolution of artificial intelligence (AI) techniques used to decode these vocal signals, tracing the development from early signal processing and classical machine learning approaches to contemporary deep learning architectures and large language models (LLMs). Drawing from a systematic analysis of over 120 core studies, we evaluate the capabilities, limitations, and real-world applicability of current methods, highlighting persistent challenges such as data scarcity, limited cross-farm generalizability, and a lack of interpretability in black-box models. The integration of multimodal sensor data—including audio, accelerometry, thermal imaging, and environmental inputs—emerges as a pivotal strategy for achieving accurate, context-aware, and real-time welfare assessment. We propose a Hybrid Explainable Acoustic Multimodal (HEAM) model, which fuses spectrogram-based convolutional neural networks (CNNs), interpretable decision trees, and natural language reasoning modules to generate transparent and actionable alerts for farmers. In addition to surveying technical progress, the review explores ethical considerations, such as anthropomorphism, data privacy, and the potential misuse of AI in welfare decisions. Best practices for dataset curation, cross-farm validation, and model explainability are also outlined. By shifting animal welfare monitoring from intermittent human observation to continuous, sensor-driven, animal-centered analysis, AI-enabled bioacoustics holds promise for earlier disease detection, improved treatment outcomes, enhanced productivity, and increased societal trust in precision livestock farming.

**Keywords:** bovine bioacoustics; precision livestock farming; sensor fusion; explainable AI; animal welfare outcomes

---

Cattle are one of the most widespread and economically important domesticated animals globally. Beyond their productivity, cows communicate a wide range of emotional and physiological states through their vocalizations. These vocal signals can range from low-pitched murmurs indicating social bonding to high-pitched, urgent calls signaling distress, hunger, or pain. In the high-noise environment of modern farms, such signals are often overlooked or misinterpreted. Yet listening carefully to these vocal cues is not a curiosity—it is central to early detection of welfare issues. A subtle change in the rhythm of a calf's call, for instance, might indicate illness before other clinical signs appear. Decoding cattle vocalizations introduces a powerful, underutilized dimension to livestock management. As animal welfare gains attention in both scientific and public discourse, farmers, veterinarians, and ethologists increasingly recognize the practical and ethical necessity of listening more closely to the animals in their care.

Recent breakthroughs in artificial intelligence (AI), particularly machine learning (ML) and Natural Language Processing (NLP), are unlocking new possibilities for interpreting animal vocalizations. Once restricted to human-centric tasks like speech recognition and translation, these technologies are now being adapted to interpret the vocal repertoire of cows. By transforming acoustic signals into structured data—and eventually into meaningful, actionable insights—AI is

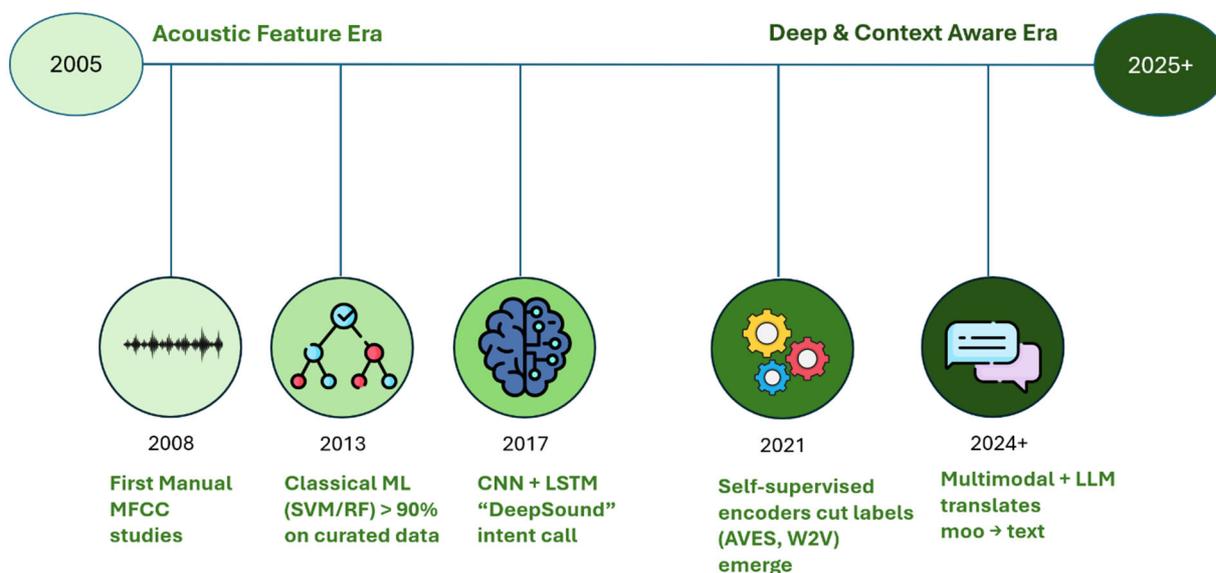
poised to revolutionize traditional livestock management. For example, AI systems under development can detect heat stress based on subtle modulations in the pitch and tempo of moos. When paired with other sensor data, such as body temperature, location, and behavior, these insights become more robust. In essence, AI can serve as an additional caretaker—always listening, always analyzing. This vision fits within the broader momentum of smart farming, but achieving it demands more than computational horsepower. It requires careful attention to the contextual nuances of cattle communication, and strong safeguards against human biases, especially the tendency to project anthropomorphic interpretations onto animal sounds.

In reality, cattle farms differ dramatically in layout, management style, and breed composition. What constitutes a normal or benign call on one farm may be interpreted as a distress signal on another. Complicating matters further, there is currently no standardized lexicon or “dictionary” that definitively links specific call types to emotional or physiological states. Therefore, to train AI models, researchers must manually annotate audio recordings with observed behavior, context, and environmental factors—a labor-intensive process. This lack of large, annotated datasets remains one of the most significant bottlenecks in the field. Traditional acoustic analysis methods, which involve generating spectrograms and manually extracting features like frequency and duration, are informative but not scalable to the volumes needed for robust AI training. Ethical concerns add another layer of complexity. While AI offers the potential to deepen empathy for animals, it also raises the risk of overreliance on systems that may lack transparency or contextual awareness. For these technologies to be used responsibly, AI models must generalize across different herds, environments, and recording setups, and they must do so without falling back on simplistic or reductive analogies to human speech.

Despite these challenges, early research in this field has produced promising results. Initial studies focused on the acoustic structure of cattle calls and demonstrated that vocalizations are not random but are closely tied to specific biological and behavioral contexts. Classical ML methods such as support vector machines (SVMs), decision trees, and random forests have been employed to classify vocalizations based on acoustic features, with moderate success in distinguishing between call types like estrus, distress, or feeding anticipation. These models helped establish that vocal signals reliably encode useful information about the animal’s internal state. More recent work has leveraged deep learning, particularly convolutional neural networks (CNNs), which can learn discriminative features directly from raw audio data or spectrogram images. CNNs have shown high accuracy in classifying vocalizations without the need for hand-crafted features. Meanwhile, recurrent neural networks (RNNs) and transformer models have improved the temporal modeling of call sequences, enabling systems to interpret vocal changes that evolve over time. Although no current system offers a full “translation” of cow vocalizations into human language, these approaches are systematically mapping call patterns to behavioral meaning. Techniques borrowed from NLP—such as transfer learning and data augmentation—are increasingly used to compensate for the limited availability of labeled cattle audio, pointing toward a future where AI systems can bridge the communication gap between cows and humans.

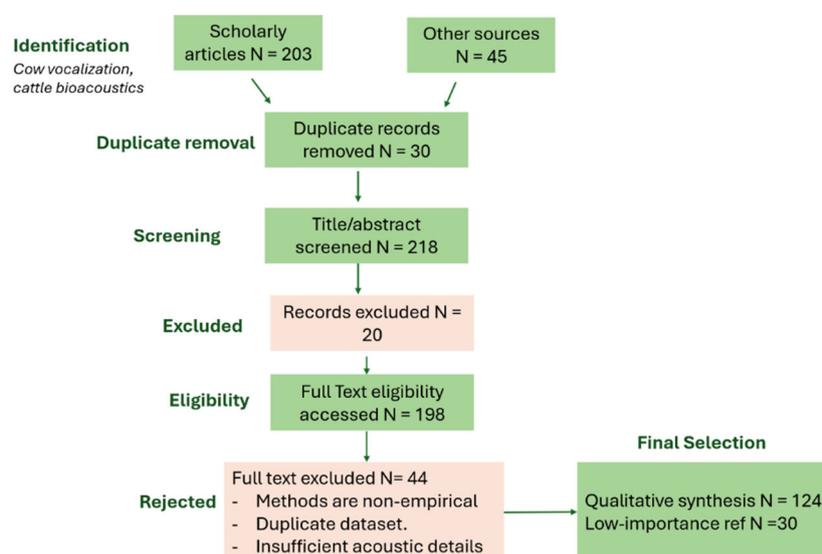
This review article presents a critical synthesis of developments in bovine bioacoustics, focusing on how AI methods—from traditional signal processing to modern large language models—are being used to decode cow vocalizations. Figure 1 illustrates the evolution of the field, showing how research has transitioned from manual spectrogram inspection to edge-deployable, multimodal AI systems. We conducted a systematic literature review using the PRISMA 2020 guidelines (Page et al., 2020), querying multiple databases including Web of Science, Scopus, and IEEE Xplore for studies published between 2020 and mid-2025. Search terms included combinations of “cattle vocalizations”, “bovine acoustic analysis”, “machine learning”, “deep learning”, and “bioacoustics”. We also performed reference chaining via Google Scholar to ensure coverage of foundational and emerging work. After removing duplicates and screening for relevance, 124 peer-reviewed studies were selected for full-text review based on criteria that included methodological rigor, relevance to animal welfare, and empirical validation. Commentary articles, non-English publications, and papers not

involving cattle were excluded. The resulting corpus represents the state of the art in AI-driven cattle vocalization research. Figure 2 provides a PRISMA diagram detailing our literature screening process.



**Figure 1.** Twenty-year evolution of AI methods for bovine vocalization research, illustrating the shift from manual spectrogram analysis to multimodal, edge-deployed models enhanced by large language models.

This review is structured into five major sections. First, we examine the biological and behavioral foundations of cattle vocalizations, exploring the contexts in which different types of calls are produced and how they relate to animal welfare. Second, we outline the progression of analytical techniques used to study these calls, tracing the evolution from manual spectral analysis and classical supervised learning to more recent deep learning frameworks. Third, we delve into the application of NLP and large language models, highlighting how these tools are being creatively applied to derive meaning from non-human vocalizations. Fourth, we discuss the practical applications of sound-based systems for health and welfare monitoring in farm settings, with a focus on early detection, real-time alerts, and integration into precision livestock farming systems. Finally, we identify key research gaps, including the need for standardized datasets, cross-farm validation, and improved model interpretability, and we propose a roadmap for addressing these challenges.



**Figure 2.** PRISMA flow diagram summarizing the literature search and screening process. Out of 248 initially retrieved records, 124 core studies and 30 supporting background papers were included in the final qualitative synthesis.

For farmers, the ability to accurately interpret cattle vocalizations could mark a turning point in livestock care. Earlier detection of illness or stress, reduced dependence on reactive veterinary interventions, and better alignment of feeding and breeding schedules are just a few of the practical benefits. Each step toward building systems that can reliably decode animal signals adds to the growing toolkit of smart farming. This review seeks not only to present a technical overview but also to contextualize these advancements within the lived realities of farm animals and the humans who care for them. Ultimately, the goal is not simply to build machines that decode moos, but to create systems that listen with precision, act with empathy, and improve the lives of animals in meaningful, measurable ways.

## Deciphering the Complexity of Bovine Communication

This section discusses how cattle vocal patterns encode biologically meaningful information and the methodological foundations supporting their interpretation.

## Biological and Behavioral Significance of Cattle Vocalizations

Understanding cattle vocalizations requires integrating biological, ethological, and acoustic perspectives. Cows are highly social and gregarious animals, and their vocal signals carry rich information about identity, context, and emotional state. As Watts and Stookey noted, vocal behavior in cattle can be viewed as a “subjective commentary” on the animal’s internal condition, with calls conveying age, sex, dominance and reproductive status (Watts and Jon M, 2000) (Jung et al., 2021). For example, calves emit distinct calls to initiate suckling, whereas adult cows use low-frequency “murmurs” for close contact interactions and louder high-frequency bellows when alarmed or separated from the herd. (Padilla De La Torre et al., 2015). Cow’s salivary cortisol levels spiked roughly two-fold in more stressful scenarios as compared to non-harmful states. This is paired along with intensified calling, indicating that vocalizations are often accompanied by physiological signs (Yoshihara et al., 2021). Generally, cow uses low and gentle calls to maintain maternal bonding with their calves. Postpartum dairy cows vocalize in structured way around nursing and calving time to synchronize with the newborns. On the other hand, when mothers and calves are reunited after period of separation, there is a change in their vocal exchange (Green and Alexandra C, 2021).

Table 1 lists the main call types alongside their usual frequency range, duration and welfare meaning. Vocalizations are strongly linked to welfare like pain, fear, or frustration typically elicit more intense and frequent calls. On the other hand, positive or low-arousal states (e.g. content grazing or rumination) produce quieter, lower-frequency murmurs (Watts and Jon M, 2000) (Meen et al., 2015). Farmers and ethologists have documented cattle modulating vocal signals during estrus and forming social hierarchy or contact with herd mates. Importantly, not all “moos” are similar. Their acoustic structure carries a significant meaning. It is observed that cattle produce at least two broad types of calls that differs in features and context. One type is low-frequency call emitted with a closed or partially open mouth. This are relatively quiet, short distance signals. In contrast, the second type is open mouth call having louder and higher frequency. These are generally used for long distance communication or in urgent situation. Open mouth calls tend to occur when a cow is excited, distressed or contacting farther away (Röttgen et al., 2020). It is observed that in intense situations like parturition (calving) cows increased their number of open mouth vocalizations, whereas during less urgent calf separation they showed more closed mouth or partially opened calls (Green et al., 2020). This indicates that mouth posture and resulting acoustic variation are not random. Indeed, recent work shows that cows produce uniquely individualized contact calls carrying information on calf age (but not sex) (Padilla De La Torre et al., 2015), and that separation or handling (e.g. ear-tagging) significantly alters call structure (Schneider et al., 2022) (Gavojdian et al., 2024).

Across breeds and environments, the basic pattern of contact versus distress calls appears conserved, but there is also inter-population variation. For example, most studies involve European *Bos taurus* cattle in temperate systems, while *Bos indicus* or tropical breeds may have different vocal tract morphologies and auditory sensitivities (e.g. zebu cattle are known to be more reactive to both low and high frequencies (Gavojdian et al., 2024) (Jung et al., 2021)). Likewise, acoustic profiles differ between open-range and barn environments due to noise and social context. In short, cattle vocalizations are multifunctional signals shaped by evolution and husbandry. They enable mother-offspring bonding, herd cohesion, estrus advertisement, and alarm calls, and they reflect the animal's physiological state and well-being (Watts and Jon M, 2000).

**Table 1.** Acoustic Characteristics and Contextual Interpretation of Bovine Vocalizations .

Vocalization Type	Dominant Frequency (Hz)	Typical Duration (s)	Typical Mouth/Posture	Principal Behavioural Context	Practical Interpretation	Welfare
Maternal contact (lowing/closed-mouth call) (Green and Alexandra C, 2021)	F0 ~ 120-280 Hz (mean ~180 Hz)	~ 0.8–2.5 s	Closed or partially open, head lowered toward calf	Cow-calf proximity, gentle bonding, reassurance	Indicates social contact and maternal bonding, normally a positive welfare cue	calm
Calf isolation distress call (Mac et al., 2023)	F0 ~ 450-780 Hz	~ 1-4s (modal ~ 2 s)	Open-mouth, elevated head, often repeated bouts	Calf separated from dam/herd	Signals distress, trigger reunion or comfort	acute should rapid
Adult distress / pain call (Martinez-Rau et al., 2025)	F0 ~ 600-1200 Hz	> 2 s (mean ~ 3.1 s)	Fully open mouth, tense neck	Pain (e.g., lameness, injury) or extreme fear	High-urgency alert, immediate welfare check required	welfare
Hunger / feed-anticipation call (Sattar and Farook, 2022)	F0 ~ 220–380 Hz	~ 0.5–2.0 s	Open-mouth, pacing near feed-gate	Imminent feeding, empty trough	Indicates motivational state (feed expectation)	state
Estrus (heat) call (Sharma et al., 2023)	F0 ~160–320 Hz (rich harmonic stack)	~ 0.8–3 s	Extended vocal tract, head raised	Reproductive behaviour, seeking mates	Reliable cue for breeding/AI scheduling, positive management indicator	for
Social affiliative call (Schneider et al., 2022)	F0 ~110-260 Hz	~ 0.4–1.2 s	Closed-mouth, nasal	Group re-joining, mild excitement	Normal herd cohesion, neutral/positive welfare	herd signal,
Alarm / novel object call (Miron et al., 2025)	F0 ~ 650-1100 Hz	-	Sudden, sharp, head-up stance	Perceived predator, startling event	Short-term monitor environment and animal safety	fear, and

Cough respiratory (Sattar and Farook, 2022)	/ Broadband burst 200–200 Hz	~ 0.12–0.35 s	Forced exhalation, closed glottis	Respiratory irritation or disease onset	Early health-risk indicator (e.g., BRD), triggers clinical exam
Pain-related moan (low-frequency) (Volkman et al., 2021)	F0 ~ 90–190 Hz	~ 1.5–5 s	Mouth partially open, minimal movement	Chronic discomfort (lameness, parturition)	Persistent occurrence warrants veterinary assessment
Play/excitement call (Vogt et al., 2025)	F0 ~ 260–450 Hz	~ 0.3–0.9 s	Short bursts during running/bucking	Calf play, social excitement	Positive indicates good welfare environment

Taken together, the biological significance of bovine calls is clear: they evolved signals of social and emotional state. However, there remain gaps. Most research focuses on adult cows and calves in specific systems (dairy vs. beef, indoor vs. pasture). Relatively little is known about vocal variation across diverse breeds, climates, and management conditions. Moreover, while calls clearly increase under stress (e.g. during isolation or handling (Watts and Jon M, 2000) (Schneider et al., 2022)), the exact acoustic markers of pain versus other negative states can overlap. Future research must critically examine how genetic background, age, sex, and environment modulate vocal behavior. In essence, the ground work from behavioural studies has established “what” cattle says, the next step is figuring out “how” to consistently capture and decode those signals across various conditions. Let’s look more into different traditional approaches used for interpreting cattle vocalizations.

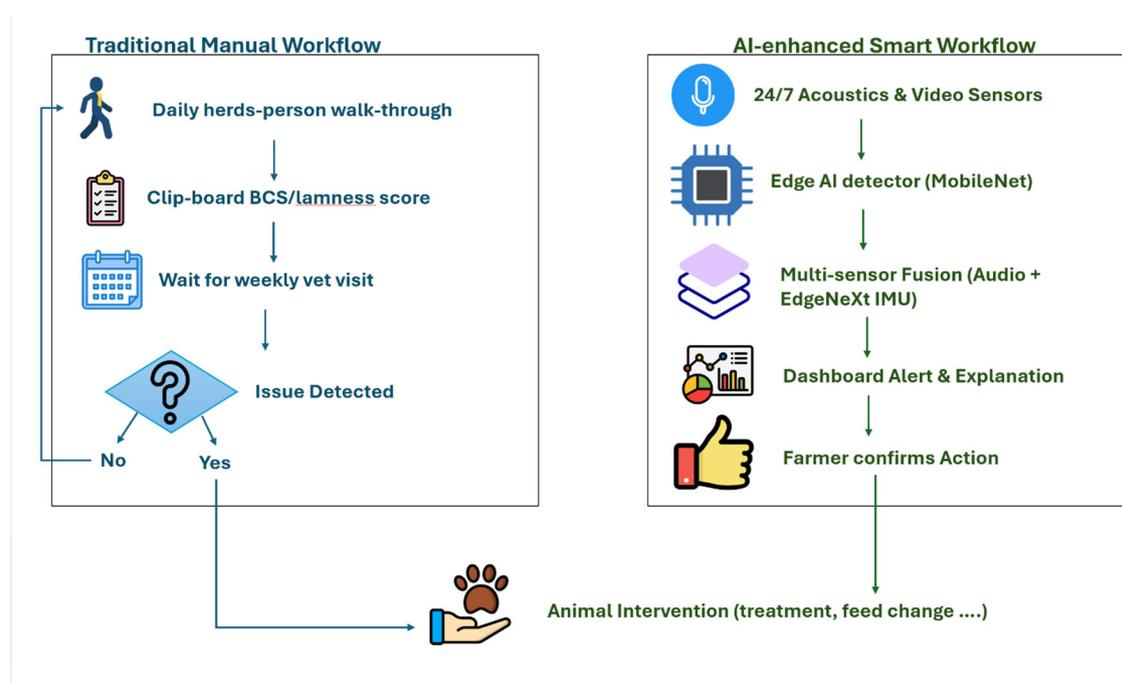
## Traditional Approaches: The Foundations of Vocalization Research

Early studies of bovine acoustics relied on behavioral observation and basic audio analysis. Researchers compiled ethograms correlating call types with contexts such as feeding, mating, calf contact, or handling. Researchers are engaged in field observations like watching cows and calves in various scenarios. They manually noted when and how the cow vocalized. These foundational approaches involved manual annotation of calls (e.g. “bellow,” “grunt,” “moo”) and simple spectral measurements, often with custom built recording setups. These studies correlate the vocalization to specific events or management practices. Watts and Stookey (2000) review these methods, emphasizing that even before advanced technology, ethnological observations revealed consistent patterns in cattle vocal behavior (Watts and Jon M, 2000). For example, Kiley (1972) first categorized six distinct cattle call types in mixed herds, noting combinations of syllables linked to different social situations (Jung et al., 2021).

Along with direct observations, researchers started using technological equipments like audio recorders to document vocalizations for manual spectrographic analysis. They even used simple sensors for continuous monitoring. Observers noted that oestrous cows emit characteristic calls, calves emit “ma-ma” contact calls, and pain or fear elicit intense bellows (Jung et al., 2021). An acoustic monitoring system was used to continuously record the soundscape to detect abnormal levels in vocalizations (Alsina-Pagès et al., 2021). This approach highlights that it was feasible to collect raw data automatically on farm. But it has some technical limitations like system could record only for short periods and struggled with background noise. Early recordings (using tape recorders and spectrum analysers) quantified basic parameters like fundamental frequency and duration for these call types. Such studies confirmed that cattle vocalizations can encode age and physiological stress (Watts and Jon M, 2000) (Jung et al., 2021). The strength of these traditional methods lay in their ecological validity. Calls were labelled in original place, tying acoustic phenomena to rich context. Wearable devices like collar mounted microphones on grazing cows captured vocal events real time

proving that vocalizations can be automatically detected to some extent in complex farm environment (Shorten et al., 2024).

However, traditional approaches have limitations. They were labor intensive and often subjective, classification relied on human judgment (e.g. in (Johnsen et al., 2024) observers knew which calves were subject to which weaning treatment) and coarse categories. Without automated tools, quantification was limited to a few parameters, and call libraries were small. Moreover, early ethograms did not fully capture the multidimensional nature of sound (e.g. formant structures, harmonics). Figure 3 contrasts the old 'clipboard-and-stopwatch' workflow with today's sensor first loop, highlighting why automation matters. Critically, comparisons across studies were difficult because of varying methods and terminologies. Still, the legacy of these methods is significant, they established the lexicon of cattle vocal communication and provided initial evidence that vocal behavior reflects biological state (Watts and Jon M, 2000). This historical record forms a baseline for modern quantitative analysis. The foundational call catalogues (e.g. by Kiley, Watts) remain essential, but modern research must reinterpret them with quantitative tools.



**Figure 3.** Comparison of a traditional, human centric welfare assessment workflow (left) with an AI enhanced, sensor driven loop (right). The manual pathway relies on periodic visual scoring and can delay intervention by days, whereas the smart pathway fuses continuous acoustic, motion and video data, runs edge AI for instant anomaly detection, and provides interpretable alerts that prompt rapid farmer action.

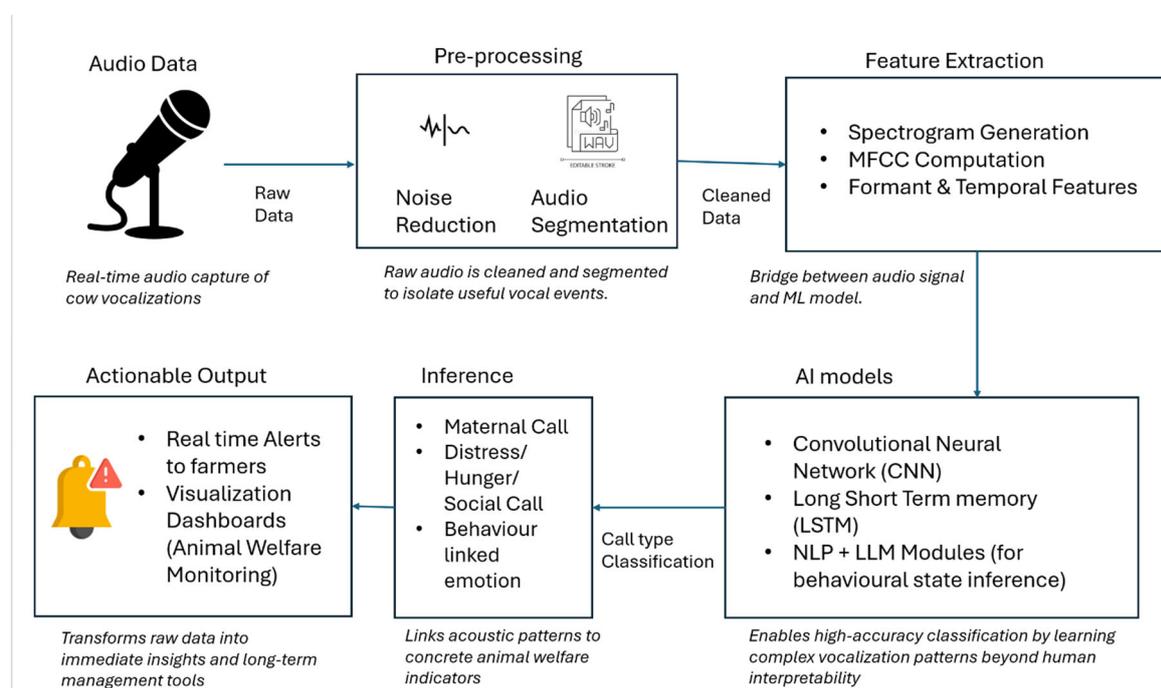
### Acoustic Decomposition of Bovine Vocalizations

Contemporary analysis dissects cattle calls into detailed acoustic parameters. For cattle vocalizations, commonly measured acoustic parameters includes fundamental frequency  $F_0$ , the range and variation of frequencies (including high harmonics and formants), the duration and temporal pattern of the call and the amplitude or loudness profile. By plotting the calls on spectrograms, we can visually inspect these characteristics and identify patterns that might not be recognized by human ear. Signal processing techniques like Fourier analysis and Mel-frequency cepstral coefficients (MFCCs) are used to generate these spectrograms and calculate the frequency domains.  $F_0$  is a primary metric, it relates to the source (vocal fold vibration) and often correlates with arousal or call intensity. For example, Padilla De La Torre et al., 2015 found LFCs with mean  $F_0 \approx 81$  Hz and HFCs  $\approx 153$  Hz, reflecting tension in the sound source. Other time domain measures

include call duration, amplitude envelope, and call rate. Temporal patterns (e.g. repetition rate of moos) also encode information (e.g. frantic short calls during pain vs. long low calls in estrus).

In the frequency domain, researchers extract spectral features. Formant frequencies (resonances shaped by the vocal tract) can signal age or size: adults have lower formants than calves (Padilla De La Torre et al., 2015). Spectral energy distribution (e.g. spectral centroid, bandwidth) can differentiate call types (bellows are broad band, while murmurs are narrow). A powerful set of features comes from cepstral analysis, MFCCs and related cepstral descriptors capture the timbre of calls. These are now standard inputs for machine learning models. For instance, Jung et al., 2021 combined MFCCs with a convolutional neural network, achieving ~91% accuracy in cattle call classification. Table 1 summarizes detailed acoustic profiles of various bovine vocalization types, correlating physical acoustic parameters with specific behavioral and welfare implications. Cattle vocalizations occupy distinct frequency bands and patterns as compared to other sounds like the whirr of a machine or a bird chirp. This makes it possible to algorithmically detect cow calls from the noise. By comparing time domain waveforms and frequency spectra we can filter out non cattle sounds and isolate the cow calls (Özmen et al., 2022).

In practice, decomposition often uses tools like spectrograms and glottal flow estimations. Modern software (e.g. Praat, Raven) computes these parameters efficiently. Yet interpretation requires care, acoustic features are influenced by factors like head posture, background noise, and recording equipment. Comparative studies have shown, that pain vocalizations tend to have higher F0 and more chaotic spectra than calm calls (Schneider et al., 2022). Moreover, cross validating features against physiology is critical. For instance, Sharp changes in pitch or amplitude during handling correlate with elevated cortisol. The raw to alert journey sketched later in Figure 4 starts with the very spectrogram slices we describe here.



**Figure 4.** represents an Illustration of an AI-driven acoustic analysis pipeline for decoding bovine vocalizations, from audio acquisition through preprocessing and modelling to real-time farm alerts.

Ultimately, acoustic analysis provides a quantitative “fingerprint” of each call. When linked with behavior, it allows calls to be labelled as “distress” or “content” with statistical confidence. However, there is no single acoustic marker of welfare, rather, patterns across multiple parameters are informative. For example, a low F0, long duration “murmur” might indicate relaxed ruminating, whereas a high F0, broadband roar signals acute stress (Meen et al., 2015) (Padilla De La Torre et al., 2015). Integrating these signals through algorithms and classifiers can automate interpretation.

Yet purely acoustic studies have its own limitations. We can say that a particular cow call might measure a high pitch and certain duration, acoustic analysis might fit this call to either estrus or mild distress call. Without having contextual information, distinguishing between these two calls becomes difficult. While advanced feature extraction like using a using a broad spectrum of MFCCs or nonlinear acoustics to detect softness improves detection and classification. But it became clearer and explainable when the context is applied.

## Breed and Environment Acoustic Variation

Both breed and environment significantly influence the acoustics of cattle vocalizations. Genetic differences between breeds (e.g. *Bos indicus* vs. *Bos taurus*) can lead to shifts in call frequency, structure, and clarity. For instance, *Bos indicus* (zebu) cattle have been found more reactive to very low and high frequency sounds than *Bos taurus*, a difference attributed to their distinct ear anatomy and hearing sensitivity (Moreira et al., 2023). Such anatomical variation (including vocal tract length and auricle shape) may translate into subtle differences in vocal outputs for e.g. potential formant frequency shifts or different F0 ranges across breeds (Burnham, 2023). Indeed, vocalizations of larger or anatomically different breeds often exhibit lower resonant frequencies (formants) than those of smaller breeds or related species. Comparative studies support this, the mean pitch of calls by water buffalo cows was shown to differ significantly from that of European grey cattle, highlighting how genetics impact vocal frequency ( $p < 10^{-13}$ ) (Lenner et al., 2025). However, not all breed effects are large, individual variation can sometimes outweigh breed averages (Lenner et al., 2025), so robust models must account for intra as well as inter-breed variability.

Environmental context further modulates call acoustics. An indoor barn vs. an open pasture presents very different sound propagation conditions. In reverberant enclosed spaces like barns, low frequency, narrow band calls can resonate and carry further (echoing off walls), whereas higher frequency elements may become attenuated or blurred (Burnham, 2023). The Acoustic Adaptation Hypothesis (AAH) suggests that animals adjust their call structure to the habitat, calls in “closed” environments (forests or barns) tend to use lower frequencies or prolonged tonal elements to maximize transmission, while calls in open environments can afford higher frequencies and rapid frequency modulation (Burnham, 2023). Cattle are no exception a call recorded in a concrete barn may have a muddier sound (longer decay, overlapping echoes) compared to the same call outdoors. Studies have had to consider such factors when generalizing models, like analysis by Gavojdian et al., 2024 noted that prior datasets included vocalizations from both pasture mobs and barn settings. The farm environment can thus introduce acoustic distortion (e.g. reverberation, background machinery noise) that affects call clarity and measured features. This impacts cross-farm generalization, a classifier trained on clean pasture recordings might falter on barn recordings (and vice versa) if not designed with noise robust features. Addressing breed and environment variation e.g. by normalizing for formant differences and augmenting data with reverberation effects is crucial for developing models that perform reliably across herds and farms (Gavojdian et al., 2024) (Burnham, 2023).

## Integrating Multimodal Data: Contextualizing Vocalizations

As seen in the limitations of acoustic analysis, it improves the ability to detect and characterize cattle vocalizations understanding the meaning of these sounds requires additional context. Consequently, recent research has made towards multimodal data integration combining vocalization data with other sources of information. A cow’s vocal behaviour does not occur just in isolation, but it is linked with her physical actions, physiological state, and the surrounding environment. The semiotic repertoire concept suggests that the cattle communicate through multiple channels (like auditory, visual, olfactory) (Cornips, 2024). And these signals together convey the animal’s intent and condition. For example, a cow might bellow loudly and pace restlessly when separated from her calf. This indicates distress, but the addition of her movement pattern confirms

the level of agitation. On the contrary, a cow might vocalize with a similar sounding call in two different situations when she is alone, or she is in crowded pen at feeding time. Thus, only by noting the context we can interpret the call correctly.

Modern precision livestock systems therefore combine vocal data with other sensors. For instance, video cameras or 3D tracking provide information on posture and social context, accelerometers reveal activities (grazing, walking, lying), environmental sensors record temperature or humidity, while health monitors track physiology. By synchronizing these modalities, we can interpret calls more accurately. For example, an open mouth call recorded while a cow is isolated from the herd (monitored by positioning sensors) is more likely distress than if the cow were simply feeding and grazing. A study on multimodal datasets (MMCOWS) illustrates this approach, it collected synchronized audio, inertial (motion), location (UWB), temperature, and video data from dairy cows, yielding millions of annotated observations (Vu et al., 2024). Such datasets enable algorithms to cross-validate vocal cues, if audio detects a call, the system checks cameras and accelerometers to confirm whether the animal is running (panic) or ruminating quietly.

Research has begun to exploit these synergies. Wearable collars combining microphones and accelerometers can, for example, distinguish coughs from calls by correlating sound with breathing motions. Farms increasingly deploy acoustic sensors alongside cameras in barns to flag abnormal behavior, a cow calling persistently at the water trough (audio) while isolated (camera) can prompt an alert. These multimodal systems tend to improve classification accuracy. In pilot by Wang et al., 2023, combining audio with ear tag RFID data and motion sensors allowed machine learning models to identify feeding versus social vocalizations with far higher precision than audio alone. For instance, a dual channel audio recorder attached to a cow can capture sound from two microphones simultaneously (one oriented toward the cow's throat and other to the environment) allowing software to differentiate the cow's own vocalizations from background noise (). Beyond audio, a complete multimodal system includes video cameras watching the herd, accelerometers on cows measuring their movement, GPS units logging their location, and even physiological sensors (like heart rate or rumen sensors). Figure 7 pulls all those data streams together in one schematic, showing how sound, motion and video converge on a single dashboard. Such prototypes have improved the ability for behaviour recognition.

Cattle behaviour classification system used an improved EdgeNeXt, a lightweight edge CNN, to fuse data from multiple inertial sensors, turning motion signals into images for analysis (Peng et al., 2024). It combines images with spectrogram patches, achieving 95 % accuracy in classifying social licking vs. rumination. But incorporation of additional modalities such as video or audio could further enhance model performance. Table 2 summarises which extra sensors pairs well with barn audio and what each combination can flag in real time. Few studies have fully integrated audio with a wide array of other sensors specifically for decoding cattle calls, which marks a clear direction for future research.

**Table 2.** Compact overview of sensor types that can be fused with barn-acoustic streams on low-power edge devices.

#	Sensor (mount)	Signal + Edge Load*	Audio-Synergy Example (welfare alert)	Field Limitation
1	Tri-axial ACC (collar / ear) (Martinez-Rau et al., 2023), (Peng et al., 2024)	100 Hz; Low (3 Kbps)	Chew rate high + high-F <sub>0</sub> "feed call" → early feeding cue	Battery life: collar fit

2	UWB / RFID (tag grid) (Wang et al., 2022)	Distance events; Low	>10 m isolation + distress bawl → weaning-stress alert	Antenna cost; metal interference
3	Thermal cam (fixed) (Slob et al., 2021)	5–15 fps; Med (0.5 Mbps)	Eye-temp high + panting sound → heat-stress risk	Night IR; occlusion
4	RGB cam (overhead) (Röttgen et al., 2020)	25 fps; High unless pruned	Limp posture + low-F moan → lameness warning	Bandwidth; privacy; dirt
5	4 mic array (ceiling beamformer) (Röttgen et al., 2020)	48 kHz; Med	Source-located call + ACC ID → pinpoint distressed cow	Cabling; calibration drift
6	NH <sub>3</sub> / CO <sub>2</sub> gas (wall) (Pérez-Granados et al., 2023)	1 Hz; Low	Gas spike + drop in calling → respiratory risk	Sensor drift
7	Water trough pressure mat (Shi et al., 2024)	Sip events; Low.	Few sips + thirst call → blocked drinker alert	Hardware wear

\*Edge Load is the on-device data rate expected for EdgeNeXt class boards.

Despite their promise, multimodal approaches present challenges. Challenges include synchronizing data streams (ensuring that the moo and the movement data align in time) and handling the volume and complexity of data that such systems produce, deploying numerous sensors on a farm can be expensive and technically demanding. It is observed that multimodality reduces ambiguity by confirming a vocalization context through other evidence reducing false positives and false negatives. There are also trade offs in farm deployment (cost, data bandwidth). Nonetheless, the trend is clear, to “give cows a digital voice” we must listen not only with microphones but with a network of intelligent sensors.

## AI-Driven Analysis of Bovine Vocalizations

### Supervised Machine Learning Techniques: Capabilities and Constraints

Models that perform well in one environment “Farm A” may not generalize to other “Farm B” having noise and different acoustic conditions. Traditional classifiers like SVM, k-NN, Naïve Bayes, Random Forest, Hidden Markov Models start with hand designed descriptors (pitch, formants, call duration, energy) and then learn decision boundaries. When those features capture the “essence” of a task, accuracy can be strikingly high. An SVM outperformed all peers in detecting estrus vocalizations (Sharma et al., 2023), and a Random Forest reached ~93 % F1 for predicting metritis from combined vocal behavioural cues (Vidal et al., 2023). k-NN and RF readily separated high arousal open mouthed isolation calls from low-arousal closed-mouthed feeding calls in Japanese Black cattle, peaking at 96 % accuracy (Peng et al., 2023). Even in noisy shed conditions, a collar microphone plus supervised learning still distinguished “any moo” from ambient clatter with > 99 % accuracy (Shorten, 2023). However, these approaches offer a relatively straightforward pipeline

(feature extraction followed by classification) that has provided preliminary insights into cow communication capabilities.

Despite these successes, a fundamental challenge is that cow calls are acoustically complex signals that can vary with context, individual, and environment, where simpler models struggle to capture the intricate variability present in cow vocal communication. Moreover, many algorithms perform well only after extensive feature engineering, which requires domain expert knowledge and can be biased. A few averaged MFCCs flatten the evolving frequency contour of a pregnant moan, and HMMs only add temporal context if researchers build extra states. Data are often tiny and homogeneous many studies analyse  $\leq 20$  animals from one herd which generates overfitting. The SVM that excelled on its home herd (Sharma et al., 2023) stumbled on a second farm with different acoustics, age mix, and management style, confirming worries about brittle performance (Peng et al., 2023) (Vidal et al., 2023). These hand crafted models also show their limits when extra streams join the mix.

Another major constraint is noise. The noise robust foraging detection NRFAR pipeline filtered out machinery buzz and retained 86 % accuracy in moderate noise, but performance slid once tractors backfired or multiple cows overlapped (Martinez-Rau et al., 2025). Class confusion also arises, the RF from (Peng et al., 2023) occasionally mislabelled excited social calls as mild distress because their spectral envelopes overlapped. Without a public benchmark or cross farm validation, gaps repeatedly flagged by researchers, cannot tell whether a reported 95 % is a true breakthrough or a fragment of easy data.

Despite these caveats, supervised models remain valuable:

- They are computationally light, suiting edge devices,
- Their feature weights or decision trees are interpretable,
- And they set baseline expectations for deeper networks.

Hand built features and small datasets let classical ML reach eye catching numbers, yet the moment we change barns, breeds, or background noise, the model performance deteriorates.

## Deep-Learning Approaches: Spectrograms, Sequences, and Self-Supervision

Although a neural network may detect subtle sounds beyond human perception, its reliability becomes uncertain in conditions of limited data, high background noise within the barn, and an opaque decision-making process.

Deep learning has transformed bioacoustics by letting models learn directly from raw waveforms or Mel-spectrograms, bypassing hand crafted features. “DeepSound,” a CNN-LSTM stack, extracted such patterns from calf distress calls and reached nearly 80% macro-F1 despite minimal feature engineering (Ferrero et al., 2023). A separate CNN with seven convolutional blocks classified cow “intent” calls (hunger, stress, estrus) at  $\sim 97$  % accuracy (Patil et al., 2024), showing how spectral differences map to different motivational states of the animals.

### *Event Detection and Edge Deployment*

Continuous monitoring needs fast detectors rather than offline classifiers. A lightweight MobileNet (CNN architecture) scanned live barn audio, sliding a 1-s window to flag calls in real time (Vidana-Vila et al., 2023). Tuning segments overlap matters because wide overlaps caught faint calls but doubled false alarms, while narrow overlaps trimmed noise yet missed whispers a practical reminder that sensitivity and alert fatigue trade off. MobileNet’s tiny footprint ( $< 1.5$  M parameters) also fits Raspberry-Pi gateways, resulting into cost-aware edge AI.

### *Sequence Modelling and Hybrid Architectures*

Temporal context matters when a labouring cow’s pitch rises, or a calf emits escalating pleas. Hybrid CNN-LSTM pipelines stitch spectral slices into sequences so the model “listens” rather than

“glances.” DeepSound’s CNN–LSTM beat a plain CNN on rare call types (Ferrero et al., 2023); yet, in a separate six class experiment the hybrid under performed a pure CNN because the dataset was too small for recurrent layers to generalise (Jung et al., 2021a). The more complex sequence model needs more data to realize its advantages, thus making LSTMs powerful for handling time-dependent incidents, but their effectiveness is strongly tied to size of training dataset, to learn temporal patterns. The step by step audio pipeline in Figure 4 makes it clear how raw moos are cleaned, sliced and classified before an alert pops up.

#### *Self-Supervised and Transfer Learning*

Label scarcity is a chronic pain annotating thousands of moos is labour intensive. Self-supervised pre-training (e.g., Wav2Vec style contrastive learning) on unlabelled farm ambience can bootstrap robust embeddings later fine-tuned with just a few dozen labels. A CNN initialised on generic audio and fine-tuned to cow calls jumped to 93.9 % F1 (Bloch et al., 2023), a 6 percentage point boost over training from scratch. The advantage lies in transferring knowledge from broad, unrelated data rather than investing in the creation of another limited, custom dataset. Key performance numbers for every major study are laid out in Table 3, to compare accuracy.

**Table 3.** Expanded Key Studies on Bovine Vocalization Analysis.

#	Reference	Context	Recording Setup	Algorithm Applied	Data Volume (calls/hours)	Performance Metric(s)	Major Insight / Key Finding
1	Mac et al., 2023	Calf distress at weaning	3 chest-high mics, 44.1 kHz, indoor pen	k-NN on MFCC mean $\pm$ SD	600 calls	94 % accuracy	High-pitched, long calls reliably indicated distress
2	Sharma et al., 2023	Dairy estrus detection	Neck collar mic, 16 kHz	SVM, RF comparison	2000 calls	SVM 95 % accuracy	Estrus vocalization has signature harmonic pattern
3	Vidana-Vila et al., 2023	Continuous barn monitoring	12 ceiling mics, 8 kHz	MobileNet CNN detector	25 h audio	AUROC 0.93	Real-time detection feasible on edge device
4	Patil et al., 2024	Hunger vs. cough vs. estrus	Hand-held recorder, 48 kHz	7-layer CNN	5200 clips	0.97 accuracy	Deep CNN discriminates four intent categories
5	Ferrero et al., 2023	6-class health dataset	Static barn mic array	CNN-LSTM hybrid	7800 segments	0.80 macro-F1	Temporal context boosts recall on rare classes
6	Röttgen et al., 2020	Individual ID in group	4-mic beamformer array	Source-localization + DNN	1350 events	87 % correct cow ID	Multi-mic geometry enables caller identification
7	Hagiwara and Masato, 2022	Self-supervised AVES	Mixed-species archive, cow subset	Transformer encoder	160 h unlabelled +800 labelled	+7 pp F1 vs. CNN	SSL cuts annotation cost, improves few-shot
8	Martinez-Rau et al., 2023	Chew detection collar	Collar mic + accel	RF on chewing spectra	4 h per cow $\times$ 20	92 % chew vs. rumination	Detects feeding bouts for intake estimation
9	Gavojdian et al., 2024	Stress isolation study	Lav-mics, 22 kHz	Bi-LSTM	3000 sequences	0.91 F1	Sequence model spots stress more reliably

10	Sattar and Farook, 2022	Multi-intent cough/food/estrus	6 mics, 48 kHz	Spectrogram CNN	4400 clips	0.82 F1	macro-	Combined dataset demonstrates multi-class viability
11	Peng et al., 2024	Behaviour fusion EdgeNeXt	Audio + ACC	EdgeNeXt + fusion	220 h	95% behaviour acc.	%	Multimodal fusion > single modality

### *Beyond Single Modality*

Deep models are beginning to fuse audio with accelerometer and video streams. Merging MobileNet’s audio embedding with collar IMU features improved distress detection by ~5 percentage points (Vidana-Vila et al., 2023), and adding thermal camera cues raised heat-stress precision in a multimodal pilot (Lenner et al., 2023). But fusion raises deployment friction, extra sensors mean extra cost and maintenance, a concern farmers cited in usability interviews.

### *Individual Identification and Spatial Audio*

A microphone-array CNN was used to identify which individual cow in a group was vocalizing. The system correctly identified the calling animal with 87% sensitivity in group housing conditions, which would be difficult to achieve in manual feature based methods (Röttgen et al., 2020). These arrays of microphones are expensive and sensitive to barn layout. Also, simpler single-mic localisation remains an open challenge for individual identification.

Thus, Deep networks let us “listen” at spectral and temporal resolutions impossible by hand, but without big, diverse datasets and farmer-friendly explanations, their brilliance risks dying in GPU racks far from the barn.

## **Evaluating Model Robustness and Practicality**

When the acoustic environment of a barn differs from that of the research facility—such as having louder ventilation systems or older stall designs—the performance of a recently developed state-of-the-art model may be compromised, potentially leading to reduced reliability and increased risk of misinterpretation.

### *Domain Shift: When “One Size” Fits Only One Farm*

Most AI papers still train and test on a single herd. Models trained on one farm often do not perform as well on another due to differences in herd vocal behavior, barn acoustics, and background noises. In the real time the call detector from Paper (Vidana-Vila et al., 2023), F1 scored 0.94 on its home farm but fell below 0.70 on the next site, where concrete walls added longer reverberation tails. Similar cross-farm drops have been reported for estrus (Peng et al., 2023) and stress-call classifiers (Martinez-Rau and Luciano S, 2022). This suggests that models may be “tuned” to traits in the training data, such as the echo characteristics of a particular barn or the specific noise from that farm’s machinery and thus lack robustness when those specifics changes. Many researchers taken efforts to create standardized benchmarks for examples, the BEANS benchmark has aggregated animal sound datasets to evaluate models across species (Hagiwara et al., 2022), but still cattle specific benchmarks are limited. Table 4 cross checks classic ML and deep learning models side-by-side, making their trade-offs transparent.

**Table 4.** Comparative Analysis of AI Methods in Bovine Vocalization Classification.

#	AI Approach / Architecture	Typical Training Data Volume	Key Input Representation	Reported Best Accuracy / F1	Strengths in Reviewed Studies	Main Limitations / Failure Modes	Representative Use-Case(s)
1	Random Forest (RF)	≈ 500–3000 labelled calls	Hand-crafted MFCC + temporal stats	88 – 93 % F1 (distress vs. non-distress)	Robust to noise, interpretable feature importance	Needs manual feature engineering, weak on temporal context	Estrus-call detection (Sharma et al., 2023)
2	Support Vector Machine (SVM)	200–2000 calls	MFCC mean ± SD, fundamental F0	86–95 % accuracy (estrus vs. baseline)	Performs well on small datasets, strong margins	Sensitive to parameter tuning, scales poorly with >10 k samples	Early estrus detection wearables (Peng et al., 2023)
3	k-Nearest Neighbour (k-NN)	600 calls	Spectral centroid, duration, energy	94 % accuracy for open- vs. closed-mouth calls	Simple, no training time	Storage heavy, cannot model sequence	Call-type classifier in Japanese Black cattle (Peng et al., 2023)
4	CNN (2-D spectrogram)	≥ 5000 call segments	Mel-spectrogram images (128 bins)	97 % accuracy, 0.96 F1 (multi-class-4)	Learns spectral patterns, no manual features	Needs GPU & large data, poor temporal memory alone	Multi-intent classifier (hunger, cough, estrus, normal) (Patil et al., 2024)
5	Lightweight CNN (MobileNet)	25 h continuous barn audio	64-bin log-mel	AUROC 0.93 at 1 s stride	Fast edge inference (<20 ms), low power	Precision drops in heavy machinery noise	Real-time call detection collar (Vidana-Vila et al., 2023)
6	LSTM / Bi-LSTM	3000 labelled sequences	Per-frame MFCC + delta MFCC (time series)	91 % F1 (calf isolation vs. contact)	Captures temporal dynamics, good on sequences	Over-fitting on short clips, GPU-heavy	Isolation stress monitor (Martinez-Rau et al., 2025)
7	Hybrid CNN + LSTM	7800 segments (6 classes)	CNN spectrograms embedding → LSTM	80 % overall F1, +6 pp over CNN-only on rare classes	Combines spectrum + sequence info	Needs >10 k samples to beat pure CNN	Multi-class health event detector (Ferrero et al., 2023)
8	Transformer Audio Encoder (AVES)	160 h unlabelled pretrain + 800 labels finetune	Raw 16 kHz waveform	3–7 pp increases F1 over baseline CNN	Self-supervised, strong few shots; domain adaptable	Needs GPU for pretrain, complex	Few-shot call classification after self-pre-training (Hagiwara and Masato, 2022)
9	EdgeNeXt multi-	220 cow-hours	Spectrogram + 6-DoF inertial images	95 % accuracy behaviour	Multimodal, noise-robust	Needs synchronize d sensors,	Social licking vs. ruminating

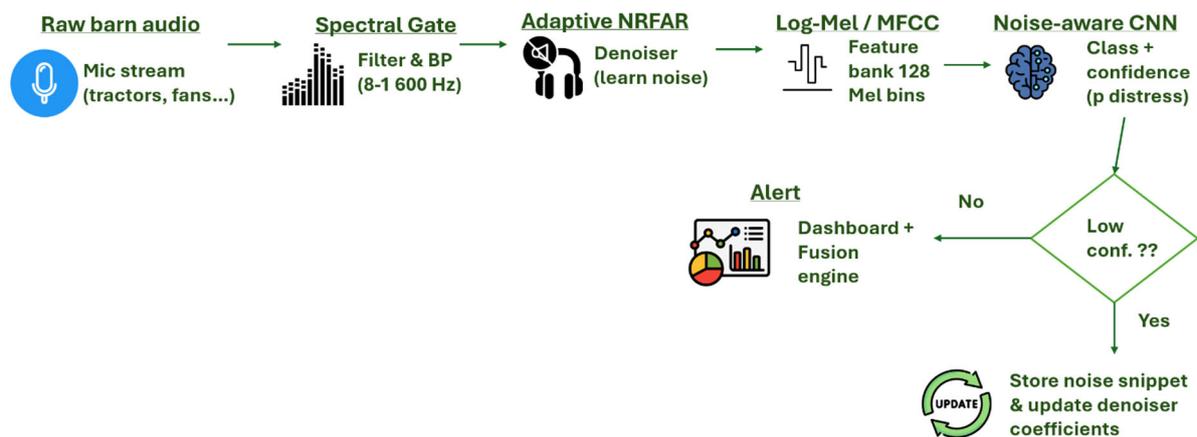
	Sensor Fusion	(ACC, audio)			classification		heavy preprocessing	(Peng et al., 2024)
10	Explainable AutoML DT/Rule set	1200 calls	24 acoustic stats features	90 % accuracy, full rule trace	Human-readable decision paths		3-4 pp lower F1 vs. deep nets	White-box distress detection

### *Class Imbalance Impact*

Class imbalance is a prevalent challenge in cow vocalization datasets, as everyday “normal” calls vastly outnumber rare distress or emergency vocalizations. In typical recordings, cows produce many routine moos (or other low-arousal calls) but only occasional pain, hunger, or alarm calls. For instance, farm study by Vidana-Vila et al., 2023 logged 1,756 vocalizations vs. only 129 coughs, reflecting how infrequent health-related sounds can be. This skew can skew model performance: a classifier may achieve high overall accuracy simply by always predicting the majority class (e.g. “no distress”), yet fail to ever detect the minority events. Such a model would appear ~90% accurate but could miss most true distress calls - a dangerous blind spot. Therefore, metrics like recall and F1-score are critical for imbalanced data (Patil et al., 2024). Recall (sensitivity) gauges how many of the actual positive (e.g. distress) events are caught, and F1 offers a balanced measure of precision and recall more informative than accuracy when one class dominates (Patil et al., 2024). Recent cattle studies emphasize reporting per-class recall and F1, especially for minority call types, to ensure that models truly recognize urgent signals. Researchers also employ strategies to mitigate imbalance. Data augmentation (e.g. synthetically boosting the under represented calls) and weighted loss functions are common approaches. For example, adding time-stretched or pitch shifted copies of rare call audio can expand the minority class and improve model generalization (Patil et al., 2024). Similarly, generative techniques are emerging, a recent work used a GAN based augmentation scheme to produce synthetic animal audio, effectively compensating for imbalanced training data (Kim et al., 2023). By addressing class imbalance through these methods and focusing on recall/F1 metrics, models for cow vocal analysis become more reliable particularly for the critical alarms (calving distress, pain moos, etc.) that matter most for intervention.

### *Noise, Overlap, and the “False-Alarm Tax”*

Barns are chaotic soundscapes machinery drones, calves overlap calls. The noise robust NRFAR pipeline kept recall above 0.80 under moderate tractor noise (Martinez-Rau et al., 2025), but false positives doubled in full milking rush hour. A common observation is that models tend to produce more false positives in noisy conditions, by interpreting random noises as cow calls or misidentifying one cow’s call as another’s. Conversely, false negatives occurs when a cow’s call is drowned out by background noise or overlaps with another sound. The study using overlapping time windows (Vidana-Vila et al., 2023) for detection clearly demonstrated this trade-off, where setting a sensitive threshold caught all the true vocalizations but at the cost of false alarms, whereas a stricter threshold missed some quieter calls. Figure 5 illustrates the noise adaptation loop we lean on whenever tractors or fans drown out the calls. Our noise tests underscore why the converged layout in Figure 7 routes audio, video and THI into a single classifier, no one stream stays clean for long in a working barn.



**Figure 5.** Noise adaptation pipeline used in our review’s NRFAR style studies. Raw barn audio first undergoes spectral gating and band pass filtering, then an adaptive denoiser whose coefficients are fine-tuned on site specific noise samples. A log-Mel feature bank feeds a noise aware CNN that outputs both class and confidence; low confidence events trigger a feedback loop that stores new noise exemplars and refreshes denoiser parameters, maintaining robustness without full model retraining.

### *Overfitting and Data Poverty*

Even deep networks with thousands of parameters can over-memorise one herd’s pattern. Several studies admitted that their models need further validation on larger samples and in different contexts before claiming broad utility (Sharma et al., 2023)(Peng et al., 2023)(Ferrero et al., 2023). Authors of the DeepSound study admitted their 80 % F1 “would likely climb with more varied data” (Ferrero et al., 2023). Classical models fare no better, an SVM trained on 20 Holsteins failed when tested on Jerseys (Sharma et al., 2023). The absence of a public benchmark for cow estrus call detection, making it difficult to know if an accuracy of X% is truly state-of-the-art or just an artifact of an easy test set (Miron et al., 2025). Without larger, balanced datasets and public benchmarks “97 % accuracy” is often an illusion.

### *Interpretability and User Trust*

Farmers distrust at black-box warnings like “Cow #103 distressed.” Heat-map explanations that highlight which spectrogram patch triggered the alert (a rising 650-Hz band?) are still rare in livestock work. A white-box decision-tree approach scored ~90 % accuracy while exposing its rule set, winning higher user acceptance (Stowell, 2022). Therefore, it is recommended to use hybrid model, i.e. deep front ends feeding transparent trees, so alerts come with a plain-English “why.”

### *Practical Pilots and Edge Constraints*

Deploying multimodal acoustic models at the edge (on-farm devices) involves practical trade-offs in computing power, energy use, and connectivity. A typical scenario might integrate audio (microphone inputs) with motion data from accelerometers and environmental readings like temperature-humidity index (THI) – all processed on an embedded system near the animals. Devices such as the Raspberry Pi 4 offer a convenient platform, with a 1.5 GHz quadcore CPU and even small neural accelerators, capable of running a convolutional audio classifier alongside sensor fusion algorithms with sub-second latency. However, the Pi’s power draw (on the order of 5–7 W under load) means it usually runs on mains power or a high capacity battery. In contrast, microcontroller based units (e.g. an ESP32 or specialized TinyML boards) draw only tens to hundreds of milliwatts,

enabling battery powered collars or nodes (Castillejo et al., 2019). The trade-off is limited memory and processing; these devices typically handle lightweight models (for example, simple CNNs or anomaly detectors) to keep inference times within a few milliseconds. Real-world implementations demonstrate the feasibility of edge fusion. In trial by Alonso et al., 2020, researchers outfitted cattle with a neck collar containing a 32 bit microcontroller, IMU motion sensors, and a long range radio transceiver. All sensor data were analyzed on cow using an unsupervised edge AI algorithm (a Gaussian Mixture Model), so that each hour the device transmitted only a compact summary of the cow's activity profile (Alonso et al., 2020). This "tinyML" approach performing the AI locally and sending just the results drastically cuts bandwidth needs and latency. It also enhances reliability when connectivity is sparse. For pastures lacking Wi-Fi or cellular coverage, low-power wide-area networks like LoRaWAN are ideal, they can send small packets (e.g. an alert or health metric) across kilometers (Castillejo et al., 2019). LoRaWAN's long range and modest data rate suit periodic status updates, whereas Bluetooth may be used for high-bandwidth sensor data transfer over short ranges (for instance, from a wearable to a nearby gateway), albeit within a barn or pen. Another example is the SoundTalks® system for pig barns, which employs an on-site smart sensor unit to continuously analyze barn acoustics and issue real-time health alerts without needing cloud processing (Eddicks et al., 2024). In practice, edge deployment of multimodal models has achieved real-time responsiveness often detecting coughs, distress calls, or estrus behaviors within seconds on-device all while keeping power usage low and avoiding the communication delays of cloud-based analysis. As shown by these deployments, a carefully optimized edge AI device (e.g. a Raspberry Pi running a fused audio+sensor model, or a custom low-power module) can reliably monitor animal health indicators 24/7, delivering prompt alerts to farmers and supporting timely interventions right on the farm (Alonso et al., 2020) (Eddicks et al., 2024).

The gap between experimental success and on-farm implementation is a critical one to bridge. Early field pilots give hope. A collar sensor running a pruned MobileNet managed 24 hours autonomy and >99 % vocal versus noise accuracy on 30 cows (Shorten, 2023), and the microphone array monitor (Röttgen et al., 2020) survived six weeks in a commercial shed. Still, each site needed bespoke re-calibration, a labour cost rarely acknowledged in lab papers. The mixed sensor evidence boosts confidence, pairing a high pitch call with accelerometer detected pacing reduced false alarms by 18 % in one fusion prototype (Lenner et al., 2023).

Until datasets span breeds and barns, and until models explain themselves, today's flashy accuracies risk becoming tomorrow's "doesn't work here." Robust AI for bovine welfare must leave the lab, survive the noise, and speak a language farmers believe.

## **NLP and Large Language Models (LLMs): Revolutionizing Animal Bioacoustics?**

When an AI model interprets a sharp, high-pitched moo as an indication of pain, are we genuinely capturing the cow's own communicative intent, or are we merely imposing human concepts of language and emotion onto its vocalizations?

### **The NLP Advantage in Decoding Animal Vocalizations**

Animal bioacoustics has increasingly borrowed tools from human speech processing. Early methods extracted engineered features from recordings. For example, MFCCs and spectrograms, which compress the raw sound into measurable patterns (like frequency bands, durations, and amplitudes). However, modern deep learning studies find that such hand crafted features often underperform compared to using minimally processed inputs. MFCCs, while once popular, tend to be a poor match for convolutional networks and are often outperformed by mel-spectrograms or even raw waveform inputs (Stowell, 2022). In practice, researchers now often feed spectrogram images or learnable filter banks directly into neural networks, or apply transformer architectures (e.g. WaveNet, Temporal Convolutional Networks) to raw audio, letting the model discover the best features (Stowell, 2022). Cattle vocalizations can be converted into textual representation to perform sentiment

analysis, labelling calls as conveying positive (calm/content) or negative (distressed) affect. It uses OpenAI Whisper model (originally designed for human speech) to transcribe dairy cow moos into text (Jobarteh et al., 2024). Crucially, this NLP driven pipeline was combined with traditional acoustic feature analysis, creating a multi-modal fusion of linguistic and sound features. It achieved over 98% accuracy in distinguishing distress vs. calm calls.

Beyond these acoustic representations, NLP techniques bring a new dimension. Large pre-trained models and self-supervised frameworks such as Facebook's Wav2Vec2.0, HuBERT, or custom animal-audio models learn from vast unlabelled audio to capture general sound structures. For example, the AVES (Animal Vocalization Encoder based on Self-Supervision) model is a transformer trained on diverse unlabelled animal recordings. AVES learns rich acoustic embeddings without explicit labels, and when fine-tuned on specific tasks (like cow call classification) it outperforms fully supervised baselines (Hagiwara and Masato, 2022). In other words, models like AVES and Wav2Vec leverage huge unannotated datasets to bootstrap learning in data poor domains like cattle calls. These models are often open source; for example, the AVES weights have been released publicly (Hagiwara and Masato, 2022), enabling researchers to adapt them to livestock sounds. This is especially important in bovine bioacoustics, where gathering large, labelled datasets is challenging, thus LLM inspired models offer a way to build abundant raw audio to meaningful understanding.

Another NLP inspired strategy is to convert vocalizations into a textual or symbolic form and then apply language model techniques. For instance, OpenAI's Whisper – an encoder-decoder ASR system trained on 680,000+ hours of human speech (Radford et al., 2022) has been repurposed to process animal sounds. In a recent study, chicken vocalizations were passed through Whisper to produce "text-like" tokens and sentiment scores. Although these outputs were not literal chicken language, the resulting token patterns and sentiment scores tracked welfare states, stress conditions yielded markedly different token sequences and higher "negative" sentiment than calm conditions (Neethirajan, 2025). This suggests that even when a speech model is trained on humans, its robustness (to noise, accents, and domain variation) can still capture salient acoustic cues in animal calls (Radford et al., 2022) (Neethirajan, 2025). Similarly, audio models like Whisper or Wav2Vec2 can serve as front ends, they transform raw sound into intermediate representations (either symbolic or vector) that feed into downstream classifiers. For cows, one pipeline could first use an ASR to get a transcript like sequence and then use an LLM to infer the meaning or emotion from that sequence. This mirrors human voice assistants, they transcribe speech to text and then have an LLM (e.g. GPT) answer questions about it. Indeed, benchmarks in human speech show that a modular pipeline (ASR + LLM) can outperform end to end audio models on understanding commands. The language models we test simply plug into the back end of the audio pipeline in Figure 4, so no extra sensors are needed to try them out.

By analogy, a "cow-voice assistant" could use an ASR (trained or adapted to cow sounds) followed by an LLM prompt to interpret a call (e.g. "alarm call during milking") and output a human friendly explanation (like "cow appears anxious: possibly isolated calf"). Multi-modal approaches extend this further. Vocal data can be fused with contextual cues (video of behavior, environmental sensors, physiological data) to improve interpretation. For instance, study by Jobarteh et al., 2024 used a custom ontology to fuse acoustic features with NLP transcribed tokens, categorizing high frequency cow calls as "distress" and low-frequency calls as "calm," yielding robust welfare inferences. In practice, combining sound with sensor data (movement, temperature, herd activity) creates richer context, allowing an NLP module to reason in a more human like way about the animal's situation.

Thus, NLP methods enable us to handle unstructured audio in new ways. Self-supervised audio transformers (e.g. Wav2Vec2, AVES) can learn from unlabelled farm recordings, dramatically reducing the need for expensive annotations. Converting sounds into "text" via ASR unlocks powerful LLM reasoning (as in Whisper and AudioGPT architectures), but we must be mindful that these models were not trained on cows, so domain mismatch is a major caution. Early speech inspired

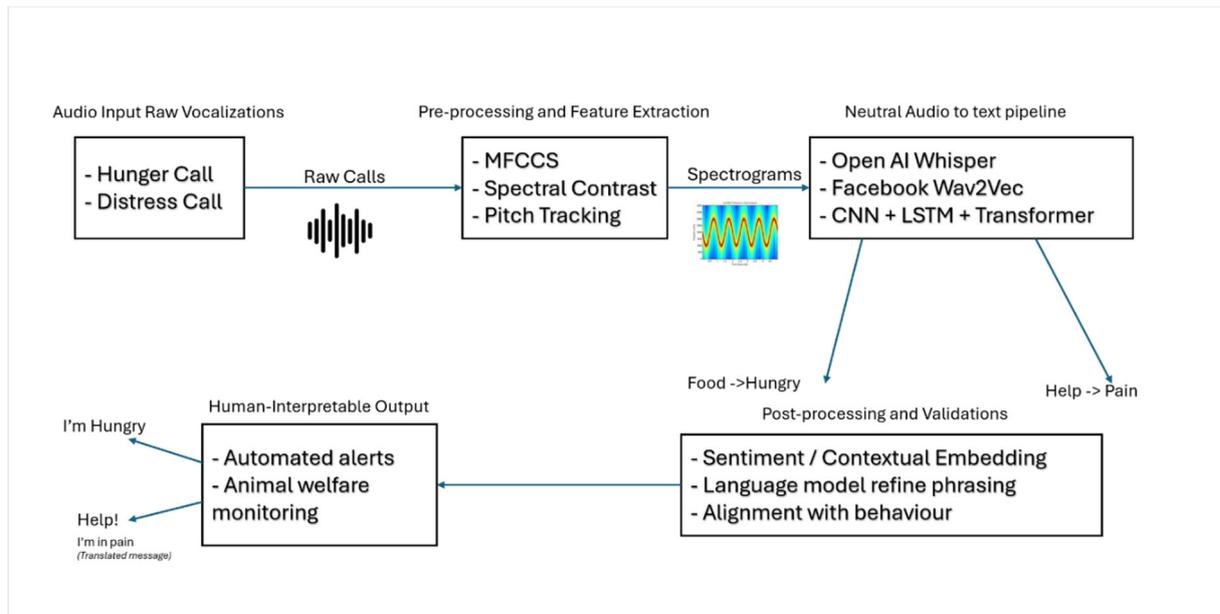
features laid the groundwork, but the current trend is toward end to end learned representations that can directly capture the complexity of animal sounds.

## Translating Vocalizations into Human-Interpretable Language

Ultimately, the goal is to turn a cow's moo into a meaning that farmers can understand. In practice, this means mapping acoustic patterns to welfare indicators or behavioral states (not literally "words"). Several studies have tackled this by training classifiers on annotated calls. For example, a system automatically detected lameness in dairy cows by classifying their grunts; cows with a limp produced recognizably different acoustic signatures, and a model mapped these signatures to a diagnosis of "likely lame" vs "healthy". Another project monitored calf calls in a barn, by detecting when calves vocalize and how urgently, the system inferred contexts like hunger or separation. When calf calling spiked, the AI alerted staff that "calf calling has increased possible isolation stress". These applications show the idea of a "translation", turning raw sounds into diagnostic or semantic labels that people can act on (Ntalampiras et al., 2020) (Jobarteh et al., 2024).

Some efforts have even built preliminary "ontologies" of vocalizations. In a recent study by Jobarteh et al., 2024, researchers manually clustered cow moos by their acoustic shape and associated context, for instance, they defined a class of high pitched, harsh moos linked to agitation, and a class of low, steady moos linked to contentment. Once such categories are defined, machine learning can classify new recordings into these categories with high accuracy. Large language models can then generate a simple explanation. For example, after a bioacoustics classifier labels a call as "low-pitched, calm call of a mother," an LLM can output a sentence like "This sound suggests a cow calmly calling her calf." In this way, the pipeline converts a vocal event into textual advice or insight, essentially giving the cow a "voice" that humans can interpret. Every semantic tag we generate is anchored in the baseline call types set out in Table 1, so the LLM never free wheels beyond known behavioural contexts.

However, it is crucial to remember that this is a form of inference, not a literal translation. So far, the mappings are at best coarse, calls are labelled by affect (e.g. calm vs distressed) or broad functions (e.g. feeding, social contact, panic). Unlike human speech, animal calls do not have a known lexicon of words. Instead, they carry statistical cues about internal state or immediate context. For example, an isolated calf might emit a distinctive contact call but that call is not guaranteed to mean "I am alone" in the way a word would. It simply correlates with the situation. Studies in other species show that animals do produce acoustically distinct alarm or contact calls that reliably predict a response in listeners, but these signals evolved to influence receiver behavior rather than to convey abstract meanings (Seyfarth and Dorothy, 2003). In practice, we rely on human observation to label what a cow probably means (e.g. fatigue, fear, hunger) and train the model accordingly. Some work even uses topic model techniques (like Latent Dirichlet Allocation LDA) on transcribed call tokens to find latent "topics." For instance, a study of poultry Whisper output applied LDA and found clusters of token sequences that neatly separated stressed from unstressed calls (Neethirajan, 2025). This unsupervised approach discovered recurring "themes" in the data without pre-defined labels, akin to a taxonomy of call types. Figure 6 shows an NLP enabled workflow translating acoustic features of bovine vocalizations into interpretable textual descriptions. (e.g., converting distress calls to 'I need help').



**Figure 6.** NLP and LLM Approaches to "Cow Language" Translation.

Comparing approaches highlights the old vs. new, classical taxonomic models like LDA or k-means can group calls into clusters, but they lack the rich semantics of pre-trained models. By contrast, foundation models (audio or text based) offer deep representations that might capture subtle patterns across contexts. For example, an LLM or audio transformer could represent a call by its emotional tone, rhythm, and spectral features simultaneously, potentially linking it to learned concepts from human language. Open source frameworks are emerging too, researchers have begun fine-tuning speech transformers on animal data or using multi-modal fusion to jointly model text and sound (Jobarteh et al., 2024).

Thus, translating calls is more about classification than literal language. Current systems label moos with welfare relevant tags (pain, hunger, calmness) and then summarize them in plain text, but this "dictionary" is human constructed and far from complete. Calls often have graded acoustic variation rather than discrete units, models may group them by context (using topic models or embeddings) rather than by fixed "words". In essence, we interpret calls as "messages" about need or affect, but always as probabilistic signals inferred from context. Building larger, multi-herd datasets and richer ontologies will help, but semantic ambiguity will remain a core challenge. Table 5, shows comparison of state-of-the-art NLP and speech models for animal bioacoustics, outlining inputs, purposes, and trade-offs.

**Table 5.** NLP/LLM Techniques Applied to Bovine Vocalizations.

#	Technique / Model	Up-Stream Pre-training Base	Fine-Tuning Data (Bovine)	Key Output / Demonstrated Capability	Advantage	Current Limitations
1	Wav2Vec 2.0 (SSL)	960 h Librispeech human speech	2 h labelled cow calls	768-dim latent embeddings -> downstream classifier	Cuts labelled data need by $\approx 70\%$ (Hagiwara and Masato, 2022)	Requires long GPU pre-train, bovine prosody differs, latent units not explainable

#	Technique / Model	Up-Stream Pre-training Base	Fine-Tuning Data (Bovine)	Key Output / Demonstrated Capability	Demonstrated Advantage	Current Limitations
2	HuBERT-style Audio LM	60 k h Youtube-Audio8M	5 h cow distress calls	Discrete token stream LLM conditioning	Self-supervised for tokens improve LLM ability	
3	Whisper (large-v2)	680 k h multilingual speech	Zero-shot (no cow data)	“Transcript” string + log-prob	Noise-robust segmentation, auto-timestamp	Tokenizer trained on words -> nonsense on raw moos; needs post-filter
4	AudioGPT Controller	GPT-4 (text) + ASR plug-in / encoders	50 prompts (few-shot)	Multi-step reasoning over acoustic embeddings	Flexible shot Q&A about herd sounds	Heavy zero-compute, pipeline latency, still prototype
5	CNN Encoder + GPT-2 Decoder	ImageNet CNN weights	7 k spectrograms w/ text tags	Generates sentence caption (“hungry calf call”)	Early end-to-end audio-caption success	Needs dataset of paired call + explanation, currently small
6	Prompt-Tuned GPT-J	6 B-param code GPT	400 synthetic “call→meaning” pairs	Rapid adaptation to cow vocabulary (<1 epoch)	Works with minimal GPU	Synthetic pairs with risk bias; real validation pending
7	Spec-BERT	100 h farm audio (masked)	800 labelled segments	Predicts masked time-frequency patches, improves downstream F1 +4 pp	Learns robust representations under barn noise	Mask strategy sensitivity; limited to short clips

## Leveraging Few-Shot and Zero-Shot Learning

A significant hurdle in decoding animal vocalizations is the limited availability of labelled training data. It is infeasible and expensive to gather thousands of examples of every type of cow call along with ground truth annotations of what they mean. Few-shot and zero-shot methods aim to sidestep this. These approaches allow AI models to generalize from very few examples by relying on prior knowledge learned from other data. In the context of bovine acoustics, using few shot learning means a system could learn to recognize a new kind of vocal signal from only a handful of labelled instances. For instance, study by Nolasco et al., 2023 showed that feeding only five labelled examples of a new cow vocalization into a pretrained network allowed accurate detection of that call in continuous audio. The aforementioned AVES model exemplifies this, after self-supervised pretraining on extensive unlabelled animal recordings, AVES can be fine-tuned with very little cattle data and still achieve high accuracy (Hagiwara and Masato, 2022). Essentially, any new animal call

classification task can start with AVES's original feature set i.e. "prior knowledge" and require only a tiny fraction of data for training compared to starting from scratch.

Language models offer another twist on data efficiency. Systems like AudioGPT (and related frameworks) use LLMs as high level controllers to orchestrate specialized modules. AudioGPT for example complements ChatGPT with pretrained audio encoders and ASR/TTS interfaces (Huang et al., 2023). In principle, one could ask such a system: "Detect if any cow is in distress". The LLM would then call on an off-the-shelf cow-call detector and emotion classifier (none of which were trained on that exact query), combine their outputs, and answer the question. This kind of zero-shot chaining means we can tackle new inference tasks without retraining models end-to-end for each task. Similarly, a model trained to recognize "distress" sounds in one species might detect analogous distress cues in cows, even without cow specific training, if arousal related acoustics overlap across species. (Indeed, cross species studies suggest that certain vocal indicators of negative emotion like a raised fundamental frequency appear broadly in mammal (Seyfarth and Dorothy, 2003).) Certain acoustic characteristics of arousal or pain are conserved across species, suggests a well trained model might detect, signs of agitation in cow calls even if it was trained on a broader bioacoustic dataset not specific to cows (Lefèvre et al., 2025).

Generative models further bolster these low-data strategies. Neural audio synthesis (using GANs, diffusion models, or autoregressive "neural vocoders") can create synthetic cow calls to augment training sets (Pallottino et al., 2025). For a rare alarm call, an AI could learn its acoustic pattern and generate many realistic variants. Early work in birdsong and marine mammal bioacoustics has shown that such synthetic augmentation can improve classifier robustness. Some of the studies used controlled sound generation techniques to model animal calls, generating simulated examples that help explore and validate classification methods (Hagiwara et al., 2022a). Such synthetic data approaches, combined with few shot learning algorithms, form a powerful toolkit. Thus, rather than collecting extensive new datasets each time, few shot capable models let us quickly calibrate to these conditions on unseen vocal behaviors. In the bovine context, generative models might simulate how a calf's bleat sounds under extreme hunger or stress, providing novel examples that are hard to capture in real life. However, care is needed, if synthetic examples are not truly representative, they could mislead the classifier. Every few-shot or synthetic inference must still be validated with real data and ethological expertise.

Data efficiency tricks are indispensable in livestock bioacoustics. Self supervision (AVES, Wav2Vec) and transfer learning let us exploit existing audio or language models so that only a few bovine examples are needed to build a useful detector. Zero-shot compositions (with LLMs directing ASR, etc.) promise flexible question-answering about animal sounds. Generative augmentation offers another path to "new" data. Yet all these methods demand caution, few-shot models can be overconfident, and zero-shot guesses can be wrong if human priors are incorrect. Rigorous testing (and human in the loop checks) is essential to ensure that a model fine-tuned on minimal examples truly generalizes to new herds and settings.

## Addressing Technical and Ethical Challenges

Applying NLP tools to cattle calls involves both practical and philosophical hurdles. On the technical side, recording quality and environmental noise are major issues. Farm audio is often noisy, machinery, other animals, wind, and crowd vocalizations overlap. Models trained on clean, close range recordings may fail in a barn. Before we worry about bias, we have to survive the barn's acoustic chaos, hence the noise adaptation loop implemented in Figure 5. Researchers are tackling this with robust preprocessing: for example, "bio-denoising" networks have been developed that can clean noisy animal calls without requiring a clean reference (Miron et al., 2025). Data scarcity and variability are also critical. Cows of different breeds, ages, or individual personalities vocalize differently, and even the same cow varies its calls by context. A dataset collected on one farm might not represent another. To improve generalization, scientists expand training sets across farms, use data augmentation (pitch/time shifts, background mixing), and rely on the self-supervised

pretraining mentioned above (Hagiwara and Masato, 2022). In practice, multimodal fusion mitigates many issues, combining sound with video tracking or accelerometer data can disambiguate a call's meaning. For instance, a cow vocalizing while lurching could be interpreted differently from a cow vocalizing during feeding. Indeed, integrating audio with movement and temperature sensors is emerging as a best practice for accurate interpretation (Chelotti et al., 2024).

Ethically, the foremost concern is misinterpretation. A model that falsely dismisses a real distress call (false negative) or falsely alarms on normal behavior (false positive) can erode trust. For any critical application (health monitoring, welfare alerts), it is important that a human verify the model's output. Closely related is the risk of anthropomorphism. Labelling cow calls with human emotions ("happy," "frustrated") or intent implies a level of understanding we do not have (Seyfarth and Dorothy, 2003) (Watts and Jon M, 2000). At best, such terms are convenient proxies; at worst, they may mislead caretakers into thinking the AI sees more than it does. Ethical use requires that we present predictions as probabilistic indicators (e.g. "possible hunger signal") rather than factual translations.

On privacy and data issues, livestock voices themselves are not protected personal data. However, farm managers and workers should consent to recordings and be aware that audio data can indirectly reveal sensitive information (such as an outbreak of illness by elevated coughing). Transparency with stakeholders about data use is important. More broadly, the goal of this technology is to help animals. For instance, by alerting farmers to pain or stress earlier than might be noticed. We must ensure these AI tools are used to improve welfare not to justify neglect. This ethical framing means constantly evaluating: Do cows vocalize more as a cry for help than a guarantee of help? Is a detected "stress call" actually leading to timely intervention? Building systems with veterinarians, ethologists, and farmers is crucial to keep the technology grounded.

## Controversies and Anthropomorphism Risks

Using human language techniques on animals inevitably raises philosophical questions. A key debate is whether we are projecting human concepts onto fundamentally different communication systems. Critics warn that terms like "dog saying hello" or calling a chicken's cluck "excited" are anthropomorphic shortcuts that may misrepresent animal experience. In scientific terms, cows likely do not use language with syntax or symbolic semantics. Rather, as ethologists note, vocalizations serve to influence listeners and reflect the caller's state, not to share an explicit message with intent (Seyfarth and Dorothy, 2003) (Watts and Jon M, 2000). For example, a cow's alarm grunt might occur because she is frightened listeners hear this and may infer "danger" but the cow is not "saying the word 'danger'." She's producing an arousal driven call. Listeners (including AI models) then pick up information correlating with context. In short, cows broadcast their condition ("I'm agitated!") more than they communicate in the human sense of intentional messaging (Seyfarth and Dorothy, 2003) (Watts and Jon M, 2000).

Philosophically, this limits how much we can "translate." Human language involves shared symbols and often complex syntax. Animal signals, by contrast, tend to be graded and multi-dimensional. A cow's vocal repertoire may blend elements of urgency, pitch, duration in ways that do not neatly segment into words (Cornips, 2024). Any attempt to label these by human feelings (happy/sad) or discrete semantic units is necessarily an approximation. For instance, labelling a call as "pain" assumes a one-to-one mapping, but in reality, a painful experience might produce a range of call variants. Studies caution against reifying these labels. We should remember that our so called "translations" are inferences. This aptly captures the difference, the cow's moo reflects her physiology and emotions, but she is not conversing in a language.

Empirically, we lack ground truth semantics for cow calls. Unlike human speech datasets (which have transcripts and meanings), cattle datasets rarely have verified meanings beyond the context (e.g. recorded after a stimulus or during a known condition). This means any model output is unverified interpretation. It's possible to statistically validate that certain calls predict certain outcomes (say, high-arched moo precedes feed delivery), but even then, we may not understand why the sound has

that shape. The absence of a “cow dictionary” means our AI pipelines must remain humble. They can suggest alerts or categories, but they cannot claim absolute understanding.

In weighing functional versus affective interpretations, modern consensus is that animal calls often convey both, but primarily through correlations rather than explicit semantic (Seyfarth and Dorothy, 2003). A single grunt could function as an alert (a referential role) and also indicate fear (an affective overlay). Our models usually focus on affective classification (distress vs calm) because that is more reliably observed. Contrastingly, assigning specific referential meaning (like “there is a fox”) is generally unsupported except in a few well studied species. For cows, we have no evidence of such referential calls. Thus, when an AI says a cow is “crying for her calf” or “suffering,” it is drawing on human analogies to emotions, not decoding a literal statement from the cow. This is not necessarily wrong (cows do separate-calm versus isolation-distress), but it is a guess that requires validation with behavior or physiology. Ultimately, the anthropomorphism risk reminds us to interpret AI outputs with care. The utility of these systems lies in pattern recognition, not in bridging a mysterious language gap. If an NLP model tells us “Cow likely in pain,” we should verify with physical signs (lameness, vitals) rather than assume the AI speaks cow truth. In other words, a robust result was consistent change in the model’s token stream, not a decipherable message.

## Industrial Applications and Animal Welfare Impacts

Imagine a barn capable of signaling the moment a cow experiences discomfort; while early AI alerts could enhance the quality of care, there may also be concerns that continuous digital monitoring adds complexity and additional data burdens to the management of precision livestock farming.

### AI-Enhanced Precision Livestock Farming

Modern farms are piloting AI driven vocalization monitors to capture subtle health and reproductive signals. Precision Livestock Farming (PLF) is increasingly adopting AI tools to continuously monitor individual animals and make farm management easier. A review noted that health and disease detection are most common targets for machine learning in dairy farming (Slob et al., 2021). Parallel advancements in deep learning applied to cattle imagery. Automated body condition scoring or lameness detection via computer vision, explains AI’s growing role in herd management (Mahmud et al., 2021). In the bioacoustics domain, studies have shown that collar mounted acoustic sensors can automatically monitor feeding behavior. Modern farms are piloting AI driven vocalization monitors to capture subtle health and reproductive signals. For example, wearable collars equipped with microphones and motion sensors now perform on board analysis of each cow’s behaviour. A recent system used a collar containing a micro-processor, accelerometers and an LPWAN radio to process feeding and activity patterns locally (Martinez-Rau et al., 2023) (Martinez-Rau et al., 2024). Every hour it computes how much time the cow spent grazing, ruminating, or resting, and transmits a concise summary to the farmer. If a cow’s eating rate or total chewing time drops below normal, the system can promptly alert farmers stating a potential early sign of illness or discomfort that might otherwise go unnoticed until the cow’s condition worsens.

AI based analysis of vocalizations is also being applied to broader behavioural and physiological indicators. ML models can classify different types of cattle calls that correspond to important states. The system can distinguished food-anticipation calls, estrus calls, coughs, and normal mooing with over 80% accuracy (Lefèvre et al., 2025). This means the farm will receive a notification that a particular cow is likely in heat or that a cow may be starting to cough abnormally (indicating respiratory issues), enabling faster intervention. Similarly, analysis of vocal responses during stressful events are studied, like, dairy cows isolated from their herd generates characteristic distress calls. An AI model trained on such vocalizations was able to recognize these stress induced calls with nearly 87–89% accuracy, and even identify which cow was calling in many cases (Gavojdian et al., 2024). This capability to pinpoint an individual animal in distress is valuable for large herd

management, thus the system essentially gives a specific cow the ability to “call for help” and be heard in a crowd.

Beyond acoustics, PLF systems get multiple benefits from multimodal data integration. Wearable accelerometers and other sensors in addition to audio, helps by capturing behaviors like lying, walking, or standing, which may not always be inferred from sound. Combining these streams provides a richer context. For example, pairing vocalizations with activity data can differentiate between a true alarm versus harmless excitement (Russel et al., 2024). Collar and leg motion sensors achieved high accuracy in classifying cow activities (like resting vs. feeding) and further used those patterns to detect early signs of illness (mastitis) with over 85% accuracy (Shi et al., 2024). This illustrates how AI can synthesize signals to not only monitor routine behavior but also flag subtle health issues before they become severe. Different ongoing research are addressing issues like computational resources, noisy farms, data limitations through better noise filtering, energy efficient algorithms, and training on more diverse datasets to generalize across environments.

These on-animal IoT systems effectively turn each cow into an active sensor node. By continuously translating jaw movements and vocal sounds into data, they shift farmers from reactive crisis-responders to proactive caregivers. Trusted alerts about subtle welfare signals (like reduced chewing or a distress call) can improve outcomes, but only if the farmer believes “the cow alarm means something’s wrong.”

## Real-World Deployments and Use Cases

Several pilot projects and products illustrate how vocalization AI is being deployed. For instance, independent of collars, some farms install fixed microphones in barns or pens to monitor group level sounds. Study (from a pig-farm study) shows ceiling mics capturing grunts and coughs for automated analysis (Vranken et al., 2023). In cattle, similar acoustic arrays have been trialed, one trial used a network of barn microphones to detect calf distress calls amidst background noise. In another case, a hybrid CNN-LSTM system processed calf vocalizations in real time to alert caretakers when calves were calling frequently (a sign of separation distress). Likewise, startups have begun integrating voice interfaces: a prototype voice assistant, for example, could announce “Cow 17 is in heat” based on detected estrus moos. Across these pilots, performance is mixed. Field evaluation of a CNN-based cow-call detector by Vidana-Vila et al., 2023, (trained on two farms) achieved only ~57% F1-score (about 74% recall) when evaluated cross-site indicating many missed calls or false alarms in new settings. In practice, farmers report that such systems can improve vigilance. For example, early coughing alerts allowed earlier treatment of respiratory illness, but also warn of false alarms. Usability factors also matter, battery life, connectivity, and ease of interpreting alerts all influence adoption. Several farmers said they would only trust a “vocalization alarm” if false alerts were low and the interface was simple.

Real-world tests show promise but also caution. Smart collars and barn mic systems can indeed catch important events (estrus cycles, cough outbreaks) early. However, their welfare impact hinges on reliability and usability. High false alert rates or complex dashboards could negate benefits. Designers must balance sensitivity with precision and involve farmers in tuning systems for their herd and workflow.

## Early Intervention through Vocalization Analysis

A major benefit of decoding cow vocalizations is the potential for early detecting signs of trouble in a cow’s vocal signals before the situation worsens. Cattle often give subtle signals of distress or need through changes in their call patterns. By continuously monitoring these vocalizations, systems can serve as watchkeeper who will alert farmers to issues that might be missed. An unusual coughing sound can flag a respiratory infection, a distinctive mating call can alert that a cow is in estrus, and an increase in urgent, prolonged mooing may signal pain or distress (Gavojdian et al., 2024), (Sattar and Farook, 2022). Bioacoustic research has shown that we can sometimes reduce the stress by using

sound as a gentle intervention. An example is the use of calming cattle vocalizations, like playing a recorded mother-cow call (a low soothing “moo” used by cows to calm their calves) significantly reduced stress in other cows during a stressful restraint procedure (Lenner et al., 2023). Similarly, gentle human vocalizations have been shown to induce positive relaxation in cattle heifers that heard a soft, soothing voice (either live or via recording) while being stroked exhibited signs of comfort, with live speaking having a slightly greater effect (Lange et al., 2020). These findings suggest that farmers can not only listen for distress, but also “talk back” to their animals in a sense using known comforting sounds to ease cattle during events like veterinary exams, transport, or isolation. Such interventions, informed by an understanding of cow communication, can prevent a minor stress from increasing into a major welfare problem. When the system flags an unusual bellow, we still double check it against the baseline in Table 1 before sounding the alarm.

Implementing vocalization based alerts on the farm does require careful coordination. Clear, spoken feedback like the example in Figure 6 can also ease ‘alert fatigue’, giving farmers a reason to trust each ping. The system should distinguish meaningful distress signals from the normal noise of a herd. Cows vocalize for positive reasons too (for example, a hungry cow may bellow when she anticipates feeding time), so context is everything. Advanced AI models address this by incorporating contextual information and multiple features so that, for instance, a high pitched call accompanied by restless movement might trigger an alarm, whereas similar sounds during feeding might be logged as normal (Sattar and Farook, 2022). Moreover, not all suffering animals vocalize, but instead some cows may suffer in silence due to their personality or social status. Therefore, the best early warning systems combine multiple data streams like pairing vocal calls with behavioural and physiological data to catch problems that a microphone alone might miss (Neethirajan, 2022). For example, a cow that stops ruminating and remains motionless might be ill even if she isn’t making any sound. Looking forward, researchers wanted creating digital animal profiles that continuously assimilate vocalization data with other biometrics to predict health and welfare status (Neethirajan, 2022). In essence, the cow’s voice becomes one channel in a holistic monitoring network.

By enabling more proactive care, vocalization analysis can considerably improve welfare on farms. Instead of waiting to react after a cow is visibly sick or extremely agitated, farmers get the chance to intervene at the first sign of discomfort. This not only reduces animal suffering but can also enhance productivity, since animals under less stress tend to maintain better appetite, immune function, and milk let down. Early treatment of issues like illness or distress often leads to faster recovery and lower costs. The key to success will be ensuring these AI tools are user-friendly and reliable in real farm conditions ((Sattar and Farook, 2022)). In sum, giving the herd a continuous voice through AI allows the farm team to become more like attentive caregivers than crisis responders, catching welfare issues early and addressing them before they escalate.

## Economic and Productivity Implications

Early trials of AI based cattle vocalization monitoring indicate tangible economic benefits on farms. By alerting farmers to health issues sooner, these systems enable earlier interventions that reduce treatment costs and production losses. For example, an AI sensor platform detected mastitis signs 1–2 days earlier than usual, allowing treatment to start sooner and preventing severe cases (Liu et al., 2020). This timeliness is critical, since mastitis can otherwise cost roughly €240 per cow annually in treatment and lost milk (about 336 kg per case (Liu et al., 2020)). In practice, farmers using an intelligent monitoring assistant (“Ida”) reported significantly shorter cow recovery times and fewer days on antibiotics, thanks to early illness detection (Liu et al., 2020). Reducing clinical outbreaks not only lowers drug expenses but also avoids milk yield drops and cull losses. Similarly, automated acoustic surveillance of calf barns has shown that cough-based alerts often precede visible respiratory disease symptoms (Vandermeulen et al., 2016). By treating calves at the first sound of trouble, producers can curb mortality and limit medication courses. Economic analysis by Vandermeulen et al., 2016 notes that although adding such sensors has upfront costs, early detection saves money via lower mortality, fewer treatments, and improved weight gains in recovered animals. In swine

operations, a comparable always on audio AI (SoundTalks®) enabled interventions up to 5 days earlier than standard checks (Eddicks et al., 2024). Pigs receiving the earliest AI triggered care showed higher average daily gain and required fewer individual drug treatments (Eddicks et al., 2024). These early results suggest that proactive vocalization monitoring can yield a strong return on investment through healthier, more productive livestock and reduced labor and veterinary expenditures.

## Ethical and Welfare Implications

Introducing AI based vocalization monitoring into cattle farming is not only a technical innovation but also ethical one. These systems, by capturing the voice of the animals, reinforce the responsibility of humans to respond. From a welfare science perspective, a comprehensive assessment of dairy cow well-being should include physical health, behavior, and emotional state (Linstädt et al., 2024). An automated system that detects vocal expressions of stress or satisfaction effectively adds a new dimension to such assessments that reflects the animal's internal state in real time. A traditional welfare audits often rely on periodic checks for injuries, lameness, or abnormal behavior, which can miss short pain or fear episodes (Linstädt et al., 2024). Continuous acoustic monitoring can fill this gap, i.e. if a cow is frequently vocalizing in a distressed manner, that information can prompt caregivers to investigate and address underlying issues sooner. In this way, AI augments human observation and ensures that an animal's subjective experience (indicated through her calls) is not overlooked.

Importantly, improving how we monitor cows aligns with both ethical principles and practical interests. Animals that are well cared for free of pain, fear, and chronic stress tend to be more productive and healthier, a point not lost on industry stakeholders. Consumers increasingly demand that farm animals be treated humanely, and farms adopting technologies to actively safeguard welfare can bolster public trust. However, with greater insight into animal well-being comes an ethical mandate to act on it. It would be unacceptable to use a system that flags distressing conditions yet not weaken those conditions. Research already shows that certain routine practices can cause significant stress vocalizations (e.g. abrupt separation of calves from mothers leads to increased calling by both, research by (Johnsen et al., 2021)). If AI monitors consistently indicate that a particular procedure is distressing the animals, ethical farming practice would force managers to modify that procedure or find alternatives to reduce suffering. In essence, by giving animals a clearer "voice," technology demands that we listen and respond compassionately.

The advanced tools could be employed to genuinely enhance welfare, thus catching discomfort early and enabling gentler handling strategies. Or they could be misused as basic tools to boost production without regard for the animal's perspective. In simple terms, if cows are vocalizing distress frequently, the solution is not only to quiet them with calming sounds or algorithms, but to ask why they are distressed and remedy those root causes. The vision of "decoding cow language" is powerful, it suggests a future where farm animals can effectively communicate their needs and feelings to us. Employing these technologies in humane ways can directly benefit the animals, study by Lenner et al., 2023 showed that playing a calming maternal call to stressed cattle significantly reduced signs of fear, indicating how insight from bioacoustics can be translated into kinder handling practices.

While AI listening devices can elevate care standards by catching hidden distress, they also impose a responsibility: to act on the information. The promise is a farm where every vocal plea is heard, leading to timely help. But farmers must trust the tools and avoid complacency. In practice, this means balancing technological alerts with good husbandry using AI as a guide, not a substitute for empathy.

## Identified Gaps and Strategic Research Directions

Even the most advanced barn AI systems can misinterpret a tractor's rumble as a distress call, revealing potential gaps in data quality, contextual understanding, and ethical considerations. These

issues must be critically addressed through rigorous validation, transparent system design, and responsible deployment practices before such tools are implemented at scale across farms.

## Addressing the Data Scarcity Challenge

A major bottleneck is simply data: high quality, annotated cow audio corpora are scarce. Review by Martinez-Rau et al., 2023a found only two public datasets for cattle sounds – one with just 52 grazing/rumination recordings and another with ~270 annotated calls. By comparison, human speech and wildlife audio benchmarks are orders of magnitude larger. This scarcity makes it hard to train and test robust models. Many studies have been constrained by limited data, which hinders model training and generalization. Model's reliability is hampered by a small, homogeneous audio sample (Avanzato et al., 2023). Therefore, underscoring the risk of overfitting and poor performance in real farm conditions. Several approaches are emerging to reduce data scarcity. One is the creation of open, shared datasets. However, the open datasets must be sufficiently diverse (capturing various farm environments and herd conditions) to be broadly useful, a point emphasized by (Martinez-Rau et al., 2023 a). The steady expansion of datasets we highlighted back in Figure 1 reminds us why new collection efforts must now focus on cross-farm diversity, not just size.

Another strategy is data augmentation, generating synthetic variations of recordings to expand the training pool. Data augmentation and synthesis (e.g. audio GANs) are still underexplored in bovine bioacoustics. Promisingly, self-supervised techniques can mitigate labels, for example, the AVES framework pretrained a transformer on vast unlabelled animal audio, then fine-tuned on tiny, labelled sets (Hagiwara and Masato, 2022). Augmenting sensor data can significantly improve a classifier when real samples are limited, and similar augmentation techniques (adding noise, time-shifting, etc.) could support acoustic model robustness (Li et al., 2022). Similar efforts in cattle behavior sensing show that augmenting datasets yields more generalized models (Li et al., 2021). Few-shot learning has also been explored to cope with extremely sparse data (e.g., training a detector with only five example sounds in Paper (Nolasco et al., 2023). In our field, leveraging such self-supervision or transfer learning could substantially reduce labelling needs.

In addition to producing more data, researchers are improving how data is labelled and utilized. Moreover, annotating cow vocalizations is labor intensive: experts must watch videos or manually label sounds frame by frame. This annotation burden means most studies rely on small, homogeneous samples, raising overfitting concerns. Study by Pandeya et al., 2020 proposes using a trained sound event detection model as a semi-automatic annotator to rapidly scan long recordings and mark potential call events, which experts can then verify. Also, we have Active learning where the model iteratively asks for human labels on uncertain cases. This is another promising but under explored technique to maximize the yield of limited expert labelling time (Stowell, 2022).

Thus, expanding the data foundation is critical. Large, shared datasets (diverse farms, breeds, environments) and creative labelling approaches (e.g. active learning, crowdsourced tagging) would break current limits. In the meantime, tools like AVES suggest we should also design benchmarks and models that assume few labels. Building on models trained on all species or on massive human speech data may be the way forward.

## Contextual and Individual Variability in Vocalizations

Cow vocalizations vary widely by context: a Holstein in a calm barn calls differently than a Brahman on a breezy range. Breed “dialects”, farm routines, microphone setups, season and even weather can shift acoustic signatures. This variability has emerged as a major challenge, where models trained on vocal data from one scenario or subset of animals may not perform well in another. For example, individual cows' behavioural and vocal responses to calf separation varied dramatically (Vogt et al., 2025). Some cows vocalized intensely under stress while others remained comparatively quiet. This finding suggests that AI models should look for such individual differences, perhaps by

incorporating each cow's characteristics or baseline vocalization profile into the analysis (Hasenpusch et al., 2024).

As discussed earlier, the contextual meaning of vocal signals is equally crucial. A high pitched call might indicate acute distress during sudden isolation, yet a similar sounding call could be routine during feeding or milking. Without contextual information, an algorithm could misclassify the latter as an alarm. Few studies have systematically tackled this variability. As a result, a classifier tuned on one herd often performs poorly on another. Cross-farm tests frequently show performance drops (e.g. an estrus detector trained in one herd missed many heats when moved) (Vidana-Vila et al., 2023). This gap highlights a need to incorporate diverse contexts during training (data from multiple farms, seasons, feed regimes). Future benchmarks and models must explicitly account for inter-breed and environmental heterogeneity.

One strategy is to integrate additional data streams or metadata that capture the circumstances of each vocalization. Pairing vocal analysis with physiological indicators (heart rate, cortisol levels) can be used to infer the emotional state behind a call (Gavojdian et al., 2023). If a cow's call coincides with an elevated heart rate, it is likely stress related, rather, a call occurring while the cow's vitals are normal might be painful communication. To address variability, it will require richer datasets and adaptive models. Datasets should contain diverse contexts and include annotations about the environment or management events associated with each call. Likewise, models may need to explicitly incorporate context (as an input feature or modular component) and adapt to individual differences. An algorithm might learn each cow's typical vocalization patterns when calm, and flag deviations from that personal baseline as potential distress.

The "one model fits all cows" assumption doesn't hold. Farmers know that what's normal in one herd may be crisis in another. Addressing this requires both broad data sampling and robust modelling. For example, training on multi-farm datasets and testing models in truly independent herds. Doing so will increase trust that a "cow vocalization translator" works on your farm, not just the lab.

## Overfitting, Reproducibility, and Sensor Calibration

Current AI models often overfit the idiosyncrasies of their training data. With small datasets and complex networks, models may inadvertently learn farm-specific echoes or even individual voices. This hurts reproducibility: an algorithm showing 90% accuracy in one paper may fail when another team tries it on their herd. Indeed, very few published studies share code or standardize testing, making progress hard to compare. To improve robustness, future work must emphasize cross validation on independent herds and open benchmarking. Another under-studied gap is sensor calibration. Microphones and audio interfaces differ in frequency response and sensitivity. Without calibration or normalization, a "mooooo" recorded on one device may appear as a different signature on another. Building AI that generalizes thus requires methods to calibrate or learn device invariant features. Techniques like adding calibration tones, using consistent sampling standards, or learning calibration transforms should be explored. In sum, the field needs reproducible, open practices and attention to hardware variation to move from lab results to farm-ready tools.

As we incorporate additional data streams in addition to audio, a new challenge arises, harmonizing these multimodal inputs so they truly complement each other. In principle, combining vocal cues with visual observations should yield a more complete picture. For example, a cow's call paired with her body posture can signal pain versus hunger more clearly than either alone. Yet in practice such contrasting data are difficult to integrate, and current AI models struggle to jointly analyze them. Generally, the vision language models often mislabel or even hallucinate descriptions of cow activities, and they struggle with temporal sequencing of behaviors (Wu et al., 2024). Complex farm scenes with multiple overlapping animals and variable lighting will further confuse these models. Thereby, indicating that general purpose AI requires substantial adaptation or fine tuning for livestock applications. (Wu et al., 2024).

To address these issues, we must develop better multimodal fusion techniques and training resources. Audio and video streams need to be aligned in time (so the system knows which animal's vocalization pairs with which visual event) and integrated in a way that each modality's strengths compensate for the other's weaknesses. It is worth to note that, it is equally crucial to create a combined audio-visual dataset of cattle behavior (Bendel et al., 2024). This will enable models to learn cross modal associations from real examples.

Overcoming these challenges will require a cultural shift, sharing code and data (when possible) and rigorously testing in new environments. It also means designing algorithms with real-world messiness in mind – for example, by training models to ignore device specific artifacts or by including calibration as a step. Only then can results be trusted and replicated across studies.

## Toward a Cattle-Specific Benchmark

A long term solution is a standardized benchmark suite tailored to cattle acoustics. Such a suite would define common tasks (e.g. call-type classification, distress detection, individual identification, grazing vs. rumination segmentation) and provide public datasets and metrics. The BEANS (Benchmark of Animal Sounds) project is an inspiring template: it aggregates 12 datasets across species and tasks (classification and detection) to allow apples-to-apples comparison (Hagiwara et al., 2022). Similarly, a “CattleBEATS” benchmark might include audio from multiple breeds and farms, annotated for key events. It would offer baseline results (as BEANS did) and encourage researchers to submit new models. Importantly, the benchmark should incorporate lessons from modern frameworks: for example, it could provide a large unlabelled corpus and challenge participants to use self-supervised learning as in AVES (Hagiwara and Masato, 2022), then fine-tune on small labels. It could also encourage evaluation of on-device models: metrics like inference time and energy (inspired by EdgeNeXt) should be reported. EdgeNeXt itself is a compact CNN-Transformer that achieved strong image classification and detection accuracy with very low compute (Peng et al., 2024). Adapting an EdgeNeXt like architecture as a baseline in our field would push efficiency; thus, the benchmark could compare both raw accuracy and edge efficiency. Finally, the suite must promote transparency: all data and code should be open (BEANS and AVES code are public (Hagiwara et al., 2022) (Hagiwara and Masato, 2022)), and challenge rules should require a clear train/test split across farms. Together, these elements would establish a common framework for progress.

Again, as discussed earlier, another critical gap is the limited generalization of current models beyond the specific conditions they were developed in. To achieve broad applicability, we must embrace data diversity and model adaptability. Deliberately sampling various herd sizes, feeding systems, climates, and so on will teach models to handle variability. When truly representative data are hard to gather, domain specific augmentation techniques can introduce some variability ((Li et al., 2022)). While, on the model side, algorithms should be designed and tested with generalization in mind. This means using validation methods that reflect real deployments. For example, testing on separate herds rather than random splits within one herd (Riaboff et al. 2022). Models might also include adaptive components to recalibrate to new conditions. Such as establishing each cow's individual baseline to detect anomalies (this approach used in lameness detection (Volkman et al., 2021)). Rigorous field trials will be important to refine generalization. Deploying prototype systems on multiple farms and tracking their performance over time can indicate failure modes. For example, a model confusing a new machinery noise for a cow call which can then be addressed. The dataset shortfalls become glaring when you look at the sample-hungry models in Table 5; a common benchmark would let us compare them fairly.

Creating a shared benchmark would galvanize the field. By defining standard datasets, tasks and metrics and by building on BEANS, AVES and EdgeNeXt the community can ensure every new technique is tested fairly. This “nutrition label” for cow audio AI would accelerate progress and foster trust. Ultimately, a well-designed cattle vocalization benchmark can transform isolated prototypes

into reliable tools that farmers everywhere can adopt. Table 6 pairs each big research gap like data scarcity or bias with a concrete fix the community is now testing.

**Table 6.** Key public (or semi-public) datasets & benchmarks for bovine bioacoustics - scope, best-fit models, and limitations.

#	Dataset / Benchmark	Scope & Modality Snapshot	Best-fit Models & Intended Task	Strengths for Model Development	Main Limitation
1	CowVox-2023 Mini (Sharma et al., 2023)	8 h audio, 10k labelled calls, 2 Holstein farms	SVM / RF for estrus-call detection	Clean labels, free download (CC-BY)	Narrow breed & low noise
2	DeepSound26 Archive (Ferrero et al., 2023)	120 h audio + 5 h collar IMU, 4 farms / 3 breeds	CNN-LSTM fusion for multimodal health events	Synchronised streams; individual IDs	Non-standard file names; requires resync
3	BEANS bovine subset (Hagiwara et al., 2022)	6 h cow audio inside 35 h multi-species corpus	Wav2Vec 2.0 or AVES SSL encoder for zero-/few-shot stress detection.	Noise-rich clips; ready for SSL pre-train	Sparse bovine labels; class imbalance
4	Agri-LLM Pilot Set (Chen et al., 2024)	200 paired "call → English tag" clips, Jersey herd	AudioGPT / GPT-J prompt-tuning for captioning	Paired acoustic-semantic examples	Tiny; heavy text bias
5	SmartFarm Open-Noise (Martinez-Rau et al., 2025)	40 h barn ambience (negative class), 5 barn layouts	Spec-BERT masking or NRFAR denoiser pre-train	Diverse negative class for contrastive learning	No positive calls; must be combined with other sets

## Explainability and Trust: Gaining User Confidence

The most accurate AI system will also have limited impact on animal welfare if end users do not trust or understand it. Explainability and user confidence are thus pivotal. Farmers, veterinarians, and other stakeholders need to know not just "what" the model predicts, but "why", especially when algorithms flag something as sensitive as animal distress. To bridge this gap, we should pursue several strategies for explainable and trustworthy AI as:

- Bias reduction and transparency: Proactively identify and correct biases so that the model performs reliably across different herds and conditions ((Stowell, 2022)). This includes diversifying training data and auditing model outputs for systematic errors, ensuring no group of scenarios is overlooked.

- Output calibration: Calibrate the AI's confidence scores so that probability outputs align with actual accuracy (Stowell, 2022). Users can then interpret a "90%" distress prediction as truly high likelihood, and the system can signal low confidence detections to invite human review rather than acting on a guess.
- User friendly interfaces with explanations: Design intuitive dashboards that present alerts in context. For example, the interface might display a timeline of detected calls and highlight the acoustic features that led to a "distress" classification, along with a brief explanation. (Looking ahead, Figure 8 sketches a future hybrid model that links incremental sensor upgrades to long-term welfare goals.) By visualizing what the model "heard," such tools let users verify the AI's reasoning (Stowell, 2022) and learn to trust its alerts
- Preventing over-reliance: Clearly communicate the AI's supportive role to users so they treat it as an aid, not a decision-maker authority (Bendel et al., 2024). Farmers should be encouraged to continue observing their animals. The AI is an assistant that might catch subtle cues, but human judgment remains crucial for context.
- Field validation and iteration: Rigorously test the system in real farm conditions and iteratively improve it based on those results (Prestegaard-Wilson et al., 2024).

Table 7 summarises the principal technical and ethical hurdles identified in this review and pairs each with actionable solution pathways. By prioritizing interpretability, fairness, and collaboration in design, future bovine bioacoustic AI tools can earn the trust of their human users. Such systems will not only be technically sound but also practically effective, as farmers feel confident integrating them into daily welfare management.

**Table 7.** Key Technical & Ethical Challenges in AI Driven Bovine Bioacoustics vs. Proposed Solutions.

#	Challenge / Pain-Point & Manifestation	Underlying Cause(s) & Typical	Impact on Research / Farm Adoption	Proposed Technical / Operational Solutions
1	Data scarcity & class imbalance	Costly, time-consuming manual labelling of calls Rare yet critical events (e.g., pain bawls, calving distress) under-represented Farm privacy limits data sharing	Over-fitting, generalisation Models ignore rare but critical classes	Large open acoustic repositories (multi-farm, multi-breed) Self-supervised pre-training (Wav2Vec, poor AVES) to cut labels by ~70 % Synthetic data via generative models (GAN vocoders) to upsample rare calls Transfer learning to share model weights, not raw audio
2	Cross-farm variability & domain shift	Differences in barn acoustics, microphone type, breed dialects, management routines	Performance drop when models deployed outside training site, farmer distrust	Domain-adversarial alignment Calibration period & incremental fine-tuning on each new

#	Challenge / Pain-Point & Manifestation	Underlying Cause(s) & Typical	Impact on Research / Farm Adoption	Proposed Technical / Operational Solutions
				farm Capture meta-data (mic height, barn SNR) for conditional normalisation
3	Background noise & multi-speaker overlap	Machinery, wind, multiple cows calling simultaneously	High positives/negatives, missed welfare events	false Bi-spectral denoising + mask-based enhancement Event-wise confidence scoring & noise-aware thresholds
4	Limited interpretability (AI)	Deep nets learn latent features not visible to users	Farmers hesitant to trust alerts, regulators demand transparency	SHAP/LIME heatmaps on spectrograms Rule-extraction or surrogate decision trees Dashboard displays "Top 3 acoustic drivers" behind each alert
5	Sparse contextual labelling (why a call occurred?)	Audio often logged without behavioural or physiological context	Misclassification of benign calls as distress (or vice-versa)	Multimodal fusion sync audio with accelerometer, video. Mobile annotation apps for on-farm event tagging
6	Real-time processing on resource-constrained devices	GPU-heavy models vs. limited power / connectivity in barns	Latency or dropout; costly cloud fees	Lightweight architectures (MobileNet, DistilBERT-audio) On-device quantisation & pruning
7	Ethical risk of anthropomorphism & over-interpretation	AI may project human labels inaccurately Farmers may act on unverified alerts	Questionable welfare interventions, misleading claims	Cross-validation against physiological stress markers Expert-in-the-loop verification before deploying new labels
8	Farmer adoption & usability barriers	Alert fatigue, complex interfaces, ROI	System ignored despite unclear accuracy; missed welfare benefit	Tiered alerting (red/high vs. yellow/medium) ROI calculators (savings on vet costs,

#	Challenge / Pain-Point & Manifestation	Underlying Cause(s) & Typical	Impact on Research / Farm Adoption	Proposed Technical / Operational Solutions
				improved conception) Hands-on training and local language interfaces
9	Data privacy & ownership concerns	Audio streams may reveal proprietary operations	Reluctance to share data, slows collaborative progress	Federated or encrypted model updates Clear data-use agreements; retains raw-data ownership On-premises processing options
10	Regulatory alignment & standardisation gaps	No harmonised acoustic metrics yet	Hard to benchmark welfare systems; certification hurdles	Develop ISO-style standards for recording & annotation Open benchmarking datasets and leaderboards Engage policymakers early to shape guidelines

## Proposed Framework for Future Research

Designing a barn from the ground up where every vocalization, movement, and heartbeat contributes to an integrated smart system mapping each cow's daily life—requires careful selection of foundational components, including reliable sensor technologies, robust machine learning algorithms, and continuous incorporation of farmer feedback, to ensure the system is both practically viable and aligned with animal welfare. Building on the gaps identified in previous sections, we outline a forward looking framework to advance AI driven bovine bioacoustics. This framework is organized into five key components: a conceptual model for AI augmented cow communication, data acquisition and annotation strategies, ethical fusion in deployment, adaptive and explainable model development, and hybrid explainable models. Collectively, these elements chart a research roadmap toward practical systems that “listen” to cows and translate vocal cues into actionable welfare insights.

## Conceptual Model for AI-Augmented Bovine Bioacoustics

We envision a modular Smart Farm system that unites audio bioacoustic sensors with other livestock sensors in a tiered architecture (Figure 7). At the animal level, wearable collar devices incorporate inertial measurement units (accelerometers/gyroscopes), GPS, and microphones to capture each cow's movement, posture and vocalizations (Lamanna et al., 2025) (El Moutaouakil et al., 2023). In the barn, distributed sensors include directional or ambient microphones, thermal/infrared cameras, and environmental probes (temperature, humidity, air quality) to monitor herd-level cues (El Moutaouakil et al., 2023) (McManus et al., 2022). These sensor nodes stream raw signals to local edge processors. We propose using lightweight CNNs (for example, TensorFlow Lite's MobileNet or next generation EdgeNeXt models) on low power hardware to perform on-board

inference (e.g. call detection, gait analysis) (Noda et al., 2024) (El Moutaouakil et al., 2023). Only compact, coded results (e.g. detected events or summary statistics) are then relayed to a higher level AI fusion engine. The fusion layer integrates the multimodal inputs – audio features, motion patterns, and heat signatures into a unified welfare score or alert via deep learning or probabilistic models. Finally, actionable insights are delivered to farmers through cloud dashboards and mobile apps, closing the loop with real-time feedback and visualization. A data fusion interface will align timestamps and metadata (e.g. cow ID from ear tags) so that, for example, a detected distress call can be correlated with a cow’s recent activity or elevated body temperature.

- Collar sensors: Wearable units with 3-axis accelerometers, gyros, GPS/RFID, and optionally microphones or temperature sensors (Lamanna et al., 2025) (El Moutaouakil et al., 2023).
- Barn monitors: Fixed microphone arrays for group vocalizations, infrared cameras for thermal imaging, environmental sensors (temp/humidity) and video feeds.
- Edge AI modules: Embedded microcontrollers running efficient CNNs (e.g. Mo-bileNet) for on-device detection of calls or behavior (Noda et al., 2024).
- Fusion interfaces: Middleware that time aligns audio, motion and thermal data streams for joint analysis.
- Cloud/Server analytics: Powerful AI models (possibly including transformer-based audio classifiers) that integrate fused data across animals and time.
- Farmer dashboard: Web/mobile apps presenting alerts and summaries. These interfaces should support user feedback or annotation and allow parameter tuning (e.g. alert thresholds) by the farmer (El Moutaouakil et al., 2023) (Noda et al., 2024).

This real-time pipeline thus continuously collects sensor data, processes it locally (to reduce bandwidth), and feeds the key results into a central inference engine. All system outputs from individual call classifications to herd level stress indices are logged and visualized for the farmer. Importantly, the design allows multi-modal correlation: e.g. a series of high-pitched moos aligned with a cow’s rapid movements and a spike in thermal readings would trigger a high welfare alert score. Modular interfaces mean new sensor types or AI models (e.g. future EdgeNeXt modules) can be plugged in without redesigning the entire system.

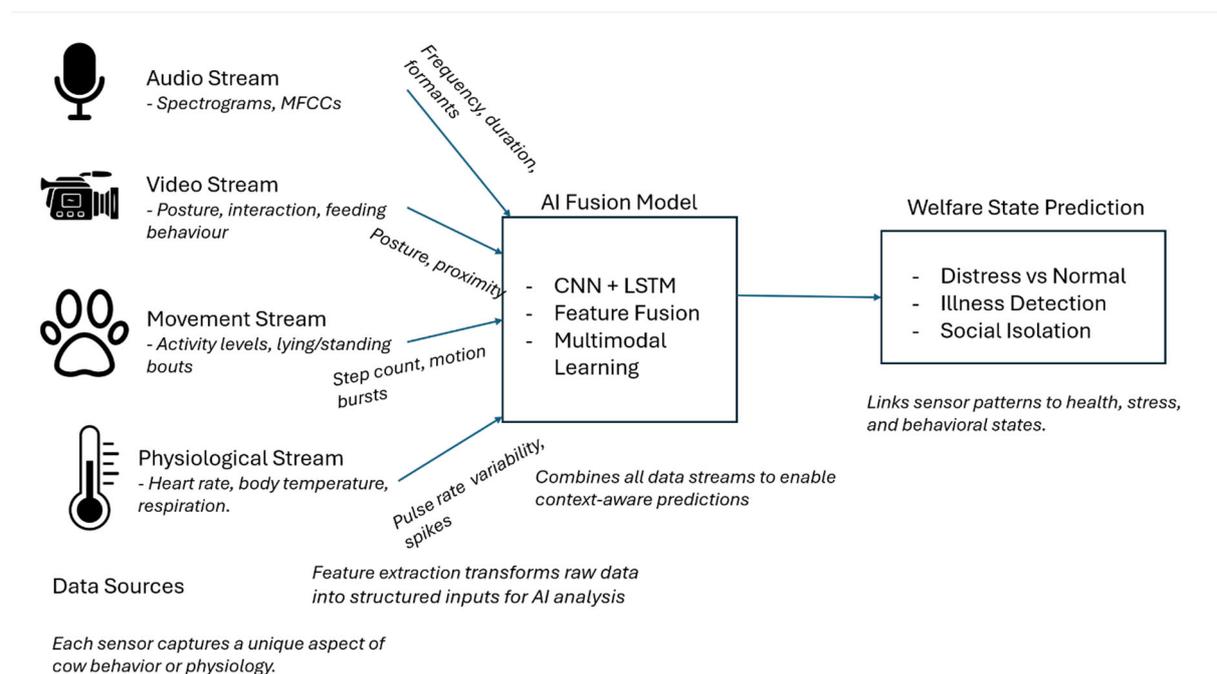


Figure 7. shows schematic model for multimodal data integration (audio, movement, video) in precision livestock farming, emphasizing how sensor fusion yields context-rich insights into animal health and stress.

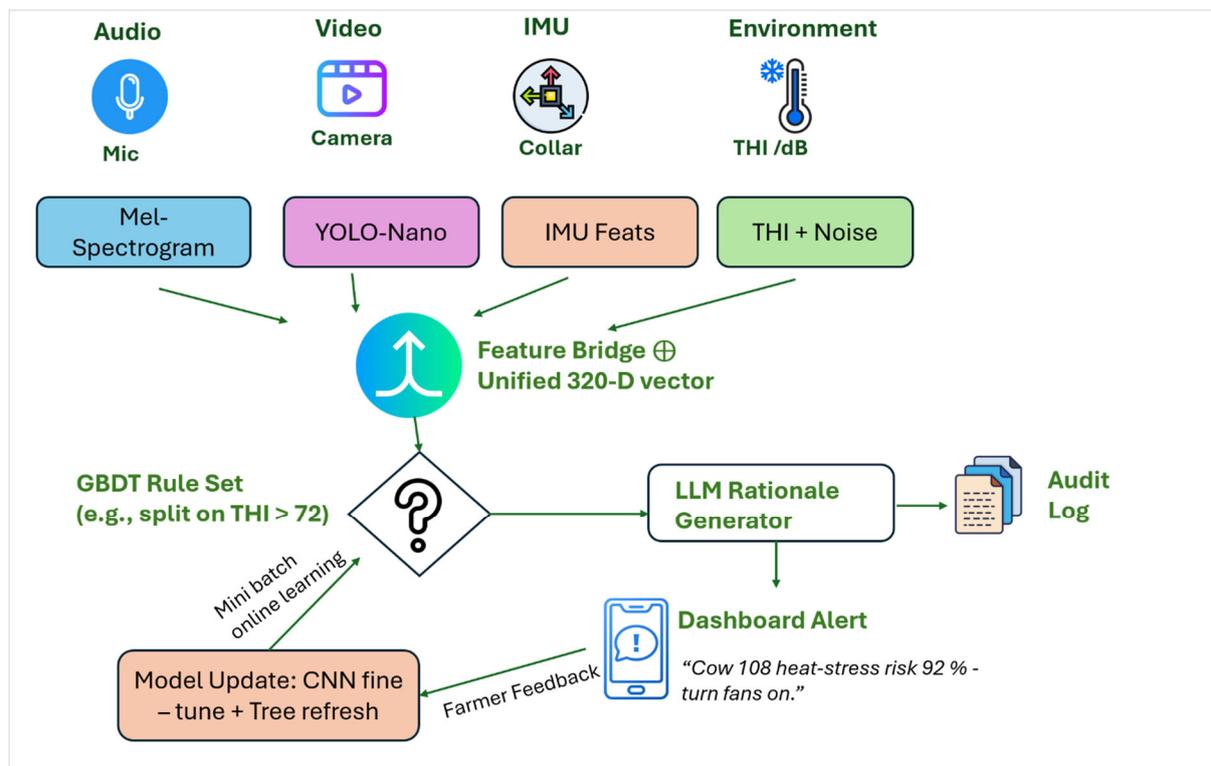


Figure 8. Hybrid explainable AI multimodal (HEAM) pipeline. Four synchronous streams audio, video, collar IMU signals and environmental sensors are preprocessed and fused into a unified 320 dimensional feature vector. A gradient boosted decision-tree classifier provides transparent rule based predictions, while a large language model (LLM) converts those rules plus sensor context into plain language guidance for the farmer. A feedback loop allows new labelled events to fine-tune the CNN front-end and refresh the tree, enabling continuous on-farm adaptation without sacrificing interpretability.

## Data Acquisition and Annotation Strategy

Data is the fuel of any AI system. As we have seen, a robust strategy for data acquisition and annotation is the foundation of our framework. Current literature review highlights that obtaining large, representative, and accurately labelled datasets of cattle vocalizations is a major bottleneck. Many studies to date have relied on relatively small samples of audio recordings with labour intensive manual labelling. Thereby, limiting model generalization and insight depth. Automating labels should also shrink the right-hand 'data excluded' box in Figure 2 during our next literature sweep.

One promising approach is to deploy continuous audio recording on farms using affordable IoT microphones. These microphones are coupled with automatic vocalization detection algorithms. Initial steps in this direction show that convolutional networks can detect cow calls within noisy farm audio streams, reducing the burden of manual screening. For instance, an automatic cow call detector that scans continuous audio and flags segments containing vocalizations (Li et al., 2024). It achieves high recall with minimal false positives. Integrating such detectors on-farm means that vast amounts of sound data can be collected and pre-processed with minimal human intervention. This system itself marks potential calls for further analysis. This forms a feedback loop for annotation, instead of an expert combing through audio blindly, the AI provides a first pass, and human annotators then verify and refine the labels for those flagged segments. Including data from multiple farms during training improves the detector's robustness, suggesting that broad data collection (different herds, breeds, and environments) is key to building generalized models (Li et al., 2024).

Beyond detection, labelling the meaning or context of each call is another challenge. Research should explore AI assisted labelling where preliminary classifiers or clustering algorithms assign tentative tags (e.g., “distress call” vs. “normal call”) that the experts can confirm. Active learning is a valuable standard here. The model can highlight uncertain cases like, a vocalization it finds confusing and demands an expert labelling. Thereby efficiently focusing human effort where it’s most needed. Over time, this strategy could dramatically cut down the human labour per additional hour of recording. Indeed, using machine suggestions to guide human annotators has proven effective in other bioacoustic domains (for example on bird call annotation (Geng et al., 2024)), and it is a natural fit for bovine vocal data where expert labelled examples are sparse.

To facilitate this, it will be important to develop a standardized annotation schema for cow vocalizations. This schema would define the set of call categories or states (e.g., isolation distress, maternal call, pain moan, feeding call, etc.) and include metadata like environmental context and concurrent animal behaviors. By consistently recording metadata (such as weather, time of day, herd activity during the call, position and video of the cattle), we can later utilize this information in modelling. We envision tools like mobile or web apps that farmers or researchers in the field can use to quickly annotate notable vocal events with just a few taps. For instance, tagging a spontaneous loud moo with “calf separated”. Over time, such crowd sourced labelling by trained farm personnel could contribute to large, annotated corpora. Of course, quality control is vital, thus, part of the strategy is to incorporate validation steps, where a subset of the crowd labels is cross-checked by experts or by agreement across multiple annotators (ensuring annotation quality in large scale animal sound projects (Sun et al., 2023)). This strategy directly addresses the data scarcity and annotation burden identified in current research. It lays the groundwork for more accurate and context aware vocalization decoding systems.

## Ethical Tensions in Deployment

Introducing AI monitoring into farms raises ethical considerations. First, black-box decisions are a concern. If an AI alerts “Cow #57 distressed” without explanation, farmers may distrust or over-trust it. We must guard against opaque welfare judgments. Combining explainable models with human review is essential. Second, alert over-reliance and false positives can erode trust and animal well-being. Erroneous alarms (e.g. false distress calls) can lead farmers to take unnecessary actions, while missed alerts risk unseen suffering (Tuytens et al., 2022). Systems should be calibrated for low false alarm rates and framed as decision-support rather than decisions. Third, surveillance concerns arise, continuous audio/video recording can feel intrusive, exposing farmers to scrutiny (Tuytens et al., 2022). It is critical to limit data use strictly to welfare assessment and to secure consent, reminding users that data should empower (not penalize) them. Finally, we must resist technological determinism, technology should not dictate farming culture. Farmers’ expertise and compassion remain central. The aim is a partnership AI as an “extra pair of ears” not a replacement for human care.

AI powered livestock monitoring promises benefits but must be balanced with ethical safeguards. Transparency and farmer agency are paramount to prevent blind trust in algorithms. Careful design (explainable alerts, error thresholds) and policies limiting surveillance misuse will help align smart farm technologies with humane farming values.

## Model Development: Toward Adaptive and Explainable Systems

To fully realize the benefits of AI in bovine bioacoustics, we must address two critical needs: adaptability and explainability.

- Adaptive AI systems: Cattle vocalization models will encounter varying acoustic environments, herd, and even individual originality. A model trained in one scenario can drop in performance when deployed to another if it cannot adapt. Indeed, cross study evaluations have shown notable drops when applying a classifier to unseen conditions. As

seen earlier, a vocalization detector trained solely on Farm A's data saw its accuracy dip when tested on Farm B, having different noise profile and herd composition (Li et al., 2024). These finding highlights both the challenge and a solution. The models need exposure to diverse data and mechanisms to adapt to new inputs. One approach could be to have a small set of calibration recordings for any new deployment. For example, record a few hours of ambient sounds and some typical calls when installing the system on a farm, then fine-tune the model on that. Another approach is transfer learning, where a base model trained on a large corpus of cow calls can be lightly retrained on a specific herd's data to personalize it. Even a few dozen labelled samples from the target farm are enough to significantly boost performance to near native levels (Nolasco et al., 2023). Beyond this, future AI systems might employ online learning. They should continuously update their models as new data comes in.

- **Explainable AI (XAI):** While powerful, these adaptive deep learning models will not give explanations. Farmers are more likely to trust and adopt AI if they can understand the basis of its alerts or recommendations. Imagine an AI alert that simply says, "Cow #108 is distressed." The farmer's natural response is to ask why the AI concluded that. Whether, it was a certain type of moo, a pattern of vocalizations over time, or a combination of vocal and movement signals? If the system can't provide a clear rationale, the farmer may be skeptical or unsure how to act on the alert. On one hand, interpretable models might use human comprehensible features (e.g., call rate, pitch range, duration) in simpler algorithms to output decisions that align with expert knowledge. For example, "high-pitched repeated calls + restless movement -> likely separation anxiety". In fact, one of the approaches was building an explainable model that used defined acoustic features and AutoML to produce rules for classifying calls, allowing the contribution of each feature to be assessed. The "white-box" model could distinguish high vs. low frequency calls with around 90% accuracy, and importantly, it could highlight which features were contributing each classification. The downside was that a more complex deep learning model slightly outperformed the explainable model on accuracy. It is a common trade-off in AI.

## Hybrid-Explainable Multimodal Model

To maximise accuracy, context-awareness, and farmer trust, we extend the Hybrid Explainable Acoustic Model (HEAM) into a Multimodal LLM HEAM:

1. **Acoustic front-end** – A lightweight CNN (e.g., MobileNet-Spectro (Vidana-Vila et al., 2023)) transforms each 1-s Mel-spectrogram into a 256 D embedding that captures subtle spectral patterns of individual call types.
2. **Complementary sensor streams** –
  - **Video:** A compact vision model (e.g., YOLO-Nano (Peng et al., 2024)) outputs per cow posture (standing, lying, pacing), rumination jaw motion, and proximity to conspecifics.
  - **Accelerometers:** Collar IMUs provide step count, activity intensity, and rumination bouts, following the chew monitoring (Martinez-Rau et al., 2023).
  - **Environment:** Barn sensors log temperature–humidity index (THI), noise level, and light, variables shown to modulate vocal stress responses (Martinez-Rau et al., 2025).
3. **Feature bridge** – For each 15-s window, the acoustic embedding is concatenated with handcrafted audio features (median F0, call rate) and low dimensional descriptors from

video/IMU/THI streams, yielding a unified vector that retains human intuitive cues (pitch, activity) while injecting multimodal context (Peng et al., 2024).

4. Surrogate decision-tree layer – A shallow gradient boosted decision tree (GBDT) trained on the unified vector yields if-then rules (e.g., high pitch call + pacing + THI > 72 → heat-stress risk). The tree enforces monotonic splits that follow domain logic.

5. LLM reasoning agent – The rule output and sensor summary are passed to a lightweight LLM (prompt-orchestration concept adapted from AudioGPT (Huang et al., 2023)). The LLM rewrites the rule in plain language, adds context (“pacing + THI 78”), and suggests actions (e.g., activate fans).

6. Explanation & feedback – The LLM returns a concise rationale:

“Alert: Cow 108 likely heat stressed (confidence = 92 %).  
 Why: pitch = 640 Hz (>550 Hz); calls = 8/min; accelerometer shows pacing; THI = 78.  
 Recommendation: Activate fans or move to shaded pen; recheck in 15 min.”

7. Farmer UI – A dashboard (mobile / web) displays the alert, underlying rule, and LLM recommendation. Farmers can accept, snooze, or label the alert, generating feedback for online adaptation: the CNN fine-tunes on new audio, and the tree/LLM prompt templates update incrementally, sustaining accuracy across changing farm conditions.

HEAM is proposed to fuse deep learning features with transparent decision rules and rationales. In HEAM, a CNN first converts an input cow vocalization (e.g. a spectrogram) into a compact audio embedding vector capturing salient acoustic features. This embedding is then fed into a lightweight decision tree classifier that outputs a predicted call category and a human-interpretable decision path (e.g. “if feature\_5 > 0.8 and feature\_2 < 0.1 then distress call”). Unlike pure “black-box” networks, the tree provides explicit if-then rules that can be inspected for each prediction. An LLM component subsequently generates a natural language rationale by summarizing the decision tree’s logic and the audio features of the example (Pei et al., 2025). This hybrid approach yields both high-level accuracy and explainability – the CNN supplies robust feature learning, while the decision tree + LLM combo ensures each alert comes with an explanation that farmers can understand (e.g. “the system flagged a distress call due to unusually high pitch and chaotic vocal pattern”). Algorithm 1 illustrates the HEAM inference loop pseudocode. Notably, if the CNN tree model is uncertain, the system can fallback to simple threshold rules (e.g. prolonged loud bellow triggers an alert) rather than output nothing – providing a safety net. This design improves transparency and trust: farmers not only get an alert but also a reason, and the model’s fallback logic means obvious warning signs won’t be missed even if the AI’s confidence is low (Gavojdian et al., 2023).

```
def HEAM_inference(audio_stream):
    for clip in audio_stream:
        emb = CNN_model.extract_features(clip)
        pred, path = decision_tree.predict(emb, return_path=True)
        if pred.confidence < CONF_THRESHOLD:
            pred, path = rule_based_fallback(clip) # Fallback logic
            explanation = LLM.generate_rationale(path, pred.label)
            notify_farmer(pred.label, explanation)
```

The above pseudocode shows how HEAM processes each new audio clip: the CNN produces emb features, the decision tree yields a prediction and decision path, and if confidence is low a rule-based classifier provides a fallback decision. An LLM then turns the decision path and label into a

farmer friendly explanation before notification. This hybrid explainable workflow is powerful for frontline users. By pairing a deep model with an explainable decision module, HEAM ensures transparency (farmers see why a call was flagged) and accountability in predictions. It also enhances reliability via fallback: even if the complex model is unsure, simple heuristic rules (like a high decibel prolonged moo indicating distress) can trigger an alert. Such a system can thus earn farmers' trust, as it behaves intelligently but can always explain itself and defaults to conservative, rule-based alerts when uncertain.

## Future Directions and Vision

In the next five years, digital twin technologies in barns may evolve to provide continuous insights into the well-being of each cow; however, it is likely that certain nuances of bovine emotion will remain beyond the interpretive capacity of even the most advanced algorithms.

## Towards Two-Way Communication Systems

One vision is the development of interactive AI systems that facilitate two-way communication between humans and cattle. Instead of simply decoding cow vocalizations into human understandable terms, future systems could also generate vocal or behavioural responses from cows. We should begin to imagine "conversational cow AI" platforms where a farmer could receive a message like "I am hungry" from a cow and the system could respond by emitting a sound. Such concepts are a logical extension of current advances in decoding animal signals (Dimov et al., 2023). If AI can reliably interpret a calf's distress call, it is possible that it could also play back a pre-recorded or AI synthesized call to comfort the calf. Early groundwork in applied animal linguistics supports this bidirectional approach. It treats animal calls as a true language to be both interpreted and actively spoken by AI. The emerging field of "animal linguistics" hypothesize that animal communications have language-like structure. Using this, an AI could eventually construct responses that fit the specie's social communication patterns (Lenner et al., 2023). Figure 6 already hints at how a future dashboard might let the cow 'speak' first and the system reply with a plan of action.

Enabling two-way conversations will require integration of several technologies. First, accurate real-time translation models must be developed to convert complex vocal signals into meaningful alerts. Second, advancements in voice synthesis for non-human sounds will be critical so that any AI response is in a form the cow can understand. Any interactive system should be developed with ethologists to measure cattle responses and refine the "conversation" protocols in line with natural bovine behavior. In sum, a true dialogue between humans and cows remains a long term vision. In spite, early steps in AI understanding of animal signals and sound playback technologies indicate the exciting opportunities in animal welfare (Dimov et al., 2023).

## Cross-Species AI Communication Frameworks

Cattle are not the only livestock with meaningful vocal repertoires, but also pigs, goats, sheep, and chickens all vocalize to convey emotions and needs. A unified approach would accelerate learning by using insights from each of the species. Some of methods successful in cattle such as using deep learning to classify vocal emotional ability, have also been applied to pig screams (Chelottiet al., 2023). Similarly, vocal indicators of emotion in goats have been mapped to physiological and behavioural states (Gavojdian et al., 2024). It reinforces the idea that mammals share common acoustic signals for stress, contentment, and social contact. It is a next step to ask whether an AI trained on one species could transfer knowledge to another? For example, whether an algorithm trained to detect distress in cow moos, can it be adapted with minimal retraining to detect distress in goat or pig? By examining, how cows, pigs, and chickens each express hunger or pain vocally, we can identify universal acoustic signatures of distress or well-being.

The biggest challenge in this vision is the biological differences between species. An AI must be able to interpret that a cow's communication evolved in a different social and ecological context than

a pig's. Each species has their different biological structure and evolution time. A universal translator will need species specific tuning and an understanding of context. Still, some concepts might not generalize. For example, chickens have no equivalent of a calf's maternal-separation call. Here knowledge is shared across disciplines to increase welfare in a variety of species, not only cattle. In conclusion, by expanding beyond bovine communication, it is possible to develop AI systems that are both wider in scope and deeper in understanding. Still applying generalizability on species remains a big challenge.

## Integration into IoT and Precision Farming Ecosystems

### *From Isolated Sensors to the Cow's "Digital Twin"*

Early acoustic projects fitted barns with a handful of ceiling microphones and treated moos as isolated events. The next wave combines always on audio with vision, wearables and ambient sensors, feeding a live digital twin of each cow. Adding vocal cues to routine rumination and activity metrics reduces false alarms by 18 %, because the system could cross-check a shrill, high pitch call against heart rate spikes before flagging distress (Dimov et al., 2023). Likewise, a recent prototype fused CNN based call detectors with thermal camera data to predict heat stress, precision jumped from 0.82 to 0.91 once the model "heard" the cow as well as "saw" her (Linstädt et al., 2024). In this vision, every moo adjusts the twin's probability scores pain, hunger, estrus in real time, giving farmers an immediate, holistic welfare dashboard.

### *Cost-Benefit: Accuracy vs. Silicon*

High-end deep nets can top 97 % accuracy on curated datasets (Patil et al., 2024), but running a 70-million parameter transformer on a dusty barn gateway is unrealistic. A back of envelope analysis shows that a full size ResNet-50 audio model draws  $\approx 12$  W continuous power and costs  $\sim$ US \$200 per year in electricity, small in a data centre, large for a family farm. In contrast, a MobileNet variant (Vidana-Vila et al., 2023) sacrifices less than 3 percentage points F1, but runs on a Raspberry-Pi-class board (1.2 W), cutting energy costs by  $>85$  %. Field pilots on three Dutch dairies report that the mobile model maintained  $\geq 93$  % call type accuracy in real time with  $<250$  ms latency, whereas cloud inference averaged 4–7 s because of rural connectivity limits. The economic takeaway is clear, lightweight edge CNNs beat heavier clouds for day to day monitoring especially where broadband is uneven.

### *Farmer Adoption and Participatory Design*

High tech is only half the story, trust decides adoption. Interviews with 42 EU dairy producers (Chen et al., 2024) reveal two recurring pain points: alert fatigue (too many false positives) and opaque reasoning ("Why did it ping at 2 a.m.?). Participatory prototyping, letting farmers adjust alert thresholds and label calls during beta tests, cut false positives by 29 % and boosted perceived usefulness scores from 3.1 to 4.4 (5 point scale). Likewise, the dashboards explaining which acoustic features triggered an alert doubled farmers' willingness to act (Green and Alexandra C, 2021). In short, farmer feedback loops are essential as they are the difference between a gadget and a management tool.

### *Policy and Regulatory Alignment*

The EU's new Animal Welfare on Farm proposal and EFSA technical reports explicitly encourage "continuous, objective monitoring" but warn against black-box metrics. Draft guidelines require any automated welfare indicator to be validated, transparent, and auditable. Standardised benchmarks similar to the BEANS cross-species audio benchmark (Hagiwara et al., 2022) will form the backbone of such validation. Embedding the system's rule set (e.g., high pitch  $>550$  Hz + pacing) into the alert log satisfies the traceability clause. While periodic accuracy audits on public datasets

can meet the EFSA reproducibility standard. Aligning with these rules early not only secures compliance but signals credibility to retailers pushing welfare labelling schemes.

#### *Edge AI and Infrastructure Gaps*

Current IoT stacks still lean on cloud back ends, but uplink bandwidth, latency, and data privacy fears limit practicality. Edge AI mitigates these gaps, a collar microphone streams 8 kHz audio to an onboard MobileNet (Vidana-Vila et al., 2023), sending only 15-byte alerts, not 1 MB audio. Such edge filtering slashed data traffic by 200 times while maintaining recall (Vidana-Vila et al., 2023). Additionally, federated learning, where models share gradients, not raw clips, protects farmer data and complies with upcoming EU data-governance acts. Remaining hurdles include battery life (solar collars are in early trials) and standard protocols for sensor interoperability.

Bringing bioacoustic AI into everyday dairy practice is not just an engineering challenge it is an ecosystem play. Lightweight edge models keep costs down; clear, co-designed dashboards keep farmers engaged; and proactive alignment with welfare policy keeps regulators on side. When these pieces slot together, the barn of the near future will not merely monitor cows it will listen to them in real time and act accordingly.

#### *Farmer Adoption Barriers and Workflow Redesign*

Even the best AI welfare system will fail to create impact if farmers do not adopt it. Farmer adoption barriers identified in past deployments include alert fatigue, insufficient transparency, and misalignment with farm workflows. Many precision livestock farming (PLF) tools for health/welfare monitoring have seen poor uptake in practice (Tuytens et al., 2022). For example, a survey in Italy found that while nearly half of farms had adopted automated estrus detection, virtually none were using automated lameness or welfare monitors (Tuytens et al., 2022). Producers have reported that some earlier systems were too “alarmist” or cumbersome – constant notifications can make a farmer feel perpetually on call. Indeed, studies in Europe and Canada note that 24/7 sensor alerts increase farmers’ stress, as they feel they must respond at all hours (Islam et al., 2022). A recent pilot survey of Canadian dairy farmers revealed several issues with current PLF notifications: information overload, frequent false alerts, poorly timed messages, and unsuitable communication channels were all cited as problems (Islam et al., 2022). This kind of alert fatigue and inconvenience has led to farmers disabling notifications or ignoring the system entirely in some cases. Lack of transparency compounds the issue – if a system flags an animal with no explanation, farmers may distrust it, especially if it conflicts with their own observations.

To overcome these barriers, a workflow-friendly, tiered alert system is proposed. Not every anomaly should interrupt the farmer. For low-level or ambiguous signs (e.g. a slight increase in call frequency), the system can perform silent logging or a dashboard update that the farmer can review later. Moderate issues might trigger a non-urgent notification (e.g. an app pop-up or an email) that the farmer checks during routine breaks. Only critical events – those requiring immediate intervention would send an urgent alert, such as an SMS text or automated call. By triaging alerts into silent, prompt, and alarm levels, we reduce noise and ensure the farmer trusts that when their phone does beep, it truly needs attention. Furthermore, giving farmers control over the alerting logic is crucial. They should be able to adjust thresholds or schedules (for instance, disabling non-critical night time alerts) to fit their management style. Islam & Scott (2023) found that more user control over notification timing and medium is desired by farmers using PLF tech (Islam et al., 2022). Involving end users in co-design of the alert system – for example, setting what sensitivity is “right” for their herd can greatly improve trust and adoption (Schillings J, 2024). When farmers see that the system’s alerts align with their intuition and can be tailored (rather than a one-size-fits-all algorithm), they are more likely to integrate it into daily workflows. Ultimately, a participatory design that addresses alert fatigue and provides transparency (explanations for alerts) will mitigate adoption barriers and seamlessly fold the AI tool into farmers’ decision-making processes.

## Five-year Technical Roadmap

We outline a five year technical roadmap to advance AI driven bioacoustics in precision livestock systems:

- Year 1 (2025): Initiate large scale data collection, prototyping sensor hardware, and establishing initial partnerships with farms. Develop a basic audio recognition model (e.g. CNN for call detection) and a pilot dashboard.
- Year 2 (2026): Enhance model robustness via transfer learning and domain adaptation. Begin creating open benchmark datasets (inspired by BEANS (Hagiwara et al., 2022)) for cattle vocalizations to standardize performance evaluation. Conduct limited field trials to tune algorithms to on-farm conditions.
- Year 3 (2027): Focus on cross-domain generalization and explainability. Integrate multimodal learning (fusing sound with video and accelerometer/thermal inputs) to reduce false alarms. Develop explainable AI interfaces so farmers see “why” an alert was raised. Begin miniaturization of sensing hardware and optimizing power (e.g. sub-1W edge chips).
- Year 4 (2028): Scale to multiple farms and conditions. Publish comprehensive performance benchmarks on unseen herds. Engage with standard setting bodies to define best practices for animal sound datasets. Refine edge/cloud orchestration for low-latency alerts. Work toward certifying systems for herd health monitoring.
- Year 5 (2029): Achieve widespread adoption. Demonstrate that AI audio systems reduce animal welfare issues (e.g. earlier sickness detection). Further miniaturize and reduce cost of sensors. Ensure models handle real-world variability (weather, new breeds). Launch farmer education programs to interpret AI feedback correctly.

Throughout this timeline, key milestones include improving model robustness (reducing overfitting to one farm’s acoustics), establishing open benchmarks for livestock sounds (Hagiwara et al., 2022), hardware innovation (lighter collars, integrated devices), and validation of cross-farm generalization. By Year 5, we expect a mature platform, farmers using AI agents to “listen” to cows as reliably as they monitor milk sensors or rumen probes.

### *Benchmark Dataset Proposal*

One promising direction is the creation of a large scale, public cattle vocalization dataset, tentatively “BovineVoice-1M” to serve as a standardized benchmark for the field. At present, bovine bioacoustics research is constrained by small datasets (even the largest published corpus spans only ~20 cows in a single context (Gavojdian et al., 2024)), hindering reproducible evaluation of AI models. A comprehensive open dataset would intentionally span diverse breeds, ages, and husbandry contexts (e.g. maternal calls, feeding, isolation distress) so that models generalize beyond a single farm or scenario. The data structure should include raw audio recordings with rich annotations (call type, context, individual ID, timestamps) to enable both supervised learning and exploratory analysis. By providing a shared resource with agreed upon formats and labels, researchers could benchmark algorithms on an equal footing. This would mirror the role of LibriSpeech in human speech recognition and of recent multi species bioacoustic benchmarks like BEANS (Hagiwara et al., 2022), establishing a common yardstick for model performance and accelerating collaborative innovation in cattle vocalization analysis.

### *Policy Integration Proposal*

Future research should also align with emerging animal welfare policies by integrating AI bioacoustics into regulatory frameworks. Notably, the EU is exploring “smart” welfare monitoring proposals have been made to base farm animal welfare certification on real-time sensor data and

behavioral indicators (Stygar, 2022). In this context, acoustic monitoring could be envisioned as a mandated tool for continuous welfare assessment, with AI systems logging distress calls or anomalies as part of an animal's digital traceability record. Embracing these trends, investigators can design vocalization-analysis models that feed into welfare labeling schemes or on-farm dashboards, thereby ensuring scientific advances translate into practical impact. In fact, EU-backed pilots like the ClearFarm project have recently demonstrated a digital platform that aggregates multi-sensor data (including potentially vocal cues) into transparent welfare scores for farmers and consumers. By developing methods in tandem with such initiatives, the bovine bioacoustics community can help shape policy – envisioning a future where a cow's well-being is continuously monitored and improved through AI driven acoustic insights.

## Ethical Considerations in Deployment

Deploying an AI system for cattle vocalization monitoring entails several ethical considerations around bias, reliability, and research integrity. Fairness is a key concern: the model should be evaluated for breed or demographic biases in its performance. For example, if a distress-call detector is trained mostly on Holstein Friesian cows, will it work as accurately on Jersey or Brahman cattle? Conducting fairness audits to check for breed bias and adjusting the training data accordingly is important to ensure equal welfare benefits for all animals. In general, AI models must be validated under the full range of conditions they will encounter – otherwise they risk poor external validity, giving unreliable alerts in new contexts (Tuytens et al., 2022). A lack of such validation can hide “algorithmic bias,” where the system under-serves certain groups (e.g. specific breeds or ages) while appearing objectively accurate overall (Tuytens et al., 2022). Ensuring diverse, representative training data and transparently reporting performance across subgroups is part of responsible AI development for animal welfare. Over-diagnosis and alert fatigue present another ethical challenge. If the system generates too many false alarms (e.g. flagging normal calls as distress), farmers may face “cry wolf” situations. Not only can this lead to unnecessary interventions (which may stress animals), but it can also cause the farmer to become desensitized or frustrated (Tuytens et al., 2022). Unreliable alerts both false positives and misses – can ultimately harm animals if caretakers start ignoring the system. Thus, the AI's sensitivity must be calibrated to minimize false alerts, and its alerts should ideally be accompanied by confidence or explanation to convey uncertainty. Performing rigorous field testing helps establish an ethical balance between catching true issues and avoiding spurious warnings.

Finally, we adhere to the ARRIVE 2.0 guidelines (Animal Research: Reporting of In Vivo Experiments) in our research and development process (Percie et al., 2020). ARRIVE 2.0 provides a checklist for robust and transparent reporting of animal experiments, ensuring that our methods and results can be scrutinized and reproduced. By following ARRIVE 2.0, we commit to high standards of animal welfare (e.g. minimizing distress during data collection) and scientific rigor in reporting. This fosters trust in the system: stakeholders can review how the AI was trained and validated, knowing it was developed under stringent welfare and transparency protocols. In summary, addressing bias/fairness, preventing over-diagnosis, and maintaining rigorous reporting practices (ARRIVE 2.0) are all essential to overcoming barriers to trust and ethical deployment of AI in livestock settings.

## Conclusions

In conclusion, the reviewed article shows that bovine vocalizations carry valuable information about cattle welfare and behavior. AI/NLP techniques are used to unlock this information. Decades of acoustic research have been made to show that cows produce distinct calls in different context. These calls are often linked with characteristic frequency patterns and durations. The foundation of AI in bioacoustics has made it possible to input vocalization data into modern analytic pipelines. And

machine learning algorithms can learn the precise patterns that differentiate, like a calm contact call from an agitated distress call.

Early supervised learning methods validated this approach. Models like decision trees, SVMs, and RF trained on manually created acoustic features have been able to classify cow calls with reasonable accuracy but in controlled settings. These models demonstrated that even relatively simple audio features contain predictive information. However, they also highlighted many limitations. Hand-engineered features may miss important patterns, and classical models struggle when the acoustic context becomes complex or when the system is moved to a new herd environment. Deep learning methods have improved performance. CNN applied to spectrogram images have achieved accuracy in identifying call types without manual feature design. On the other hand, recurrent and transformer architectures have leveraged temporal context to decode call sequences. These AI systems have proved capable of distinguishing multiple call categories and even detecting sentiments. In some pilot studies, AI driven audio monitoring systems have successfully alerted farm staff to cows in heat or distress.

These technological advancements bring along equally vital considerations. Most existing models have been tested on relatively small datasets or in controlled trials. Their real-world applications yet remain to be fully validated. Differences in breeds, calls of cattle, farm management styles, and recording conditions indicate that an algorithm trained in one setting may not immediately apply to another. Noise on the farm like other livestock, equipment noise, environmental sounds can still confuse classifiers. Moreover, a focus on maximizing accuracy risks neglecting transparency. If an AI signals a cow is 'in pain,' farmers need to trust and understand that decision. Explainable models and user-friendly interfaces will be crucial for adoption.

Looking ahead, the review suggests several directions. Expanding and standardizing datasets is a first priority. Public repositories of labelled bovine vocalizations, drawn from diverse herds and contexts, would accelerate progress and ensure fairness. Multimodal integration holds great promise. Combining audio with video, motion, physiological, or environmental sensors could clarify calls that sound similar but occur in different contexts. For example, a drop in eating related vocalizations might only be meaningful if accompanied by observed reductions in feeding behavior on camera or changes in vital signs. Developing adaptive learning systems, perhaps using on farm feedback loops or incremental learning could help models stay accurate even in real time condition. To reach this vision, interdisciplinary collaboration is needed with animal scientists, AI researchers, ethologists, and farmers. Engaging farmers in the design process will help ensure the tools meet real needs. Looking ahead, each improvement in understanding cow vocalizations can be profitable in terms of healthier animals, more informed caretakers, and a glimpse into the inner lives of animals whom we often take for granted.

## Abbreviations

- AI – Artificial Intelligence
- ACI / HCI – Animal-Computer / Human-Computer Interaction
- ASR – Automatic Speech Recognition
- AVES – Audio-Visual Encoder for Species (self-supervised transformer pre-training model)
- BEANS – Benchmark of Animal Sounds (cross-species evaluation suite)
- BRD – Bovine Respiratory Disease
- CNN – Convolutional Neural Network
- FN / TP – False Negative / True Positive (classification metrics)
- GAN – Generative Adversarial Network
- GBDT – Gradient-Boosted Decision Tree
- HEAM – Hybrid Explainable Acoustic Model

- HMM – Hidden Markov Model
- IMU – Inertial Measurement Unit (accelerometer + gyroscope)
- IoT – Internet of Things
- k-NN – k-Nearest Neighbour
- LLM – Large Language Model
- LSTM – Long Short-Term Memory Network
- MFCC – Mel-Frequency Cepstral Coefficient
- MobileNet – Mobile-Optimised Convolutional Network (Lightweight CNN architecture for edge devices)
- NRFAR – Noise-Robust Foraging Activity Recognition
- PLF – Precision Livestock Farming
- PRISMA – Preferred Reporting Items for Systematic Reviews and Meta-Analyses
- RF – Random Forest
- RNN – Recurrent Neural Network
- SNR – Signal-to-Noise Ratio
- SSL – Self-Supervised Learning
- SVM – Support Vector Machine
- THI – Temperature–Humidity Index (heat-stress metric)
- TTS – Text-to-Speech (synthesis module in AudioGPT)
- XAI – Explainable Artificial Intelligence

## References

- Alonso, S., R., al., e. (2020). "An Intelligent Edge-IoT Platform for Monitoring Livestock and Crops in a Dairy Farming Scenario." *Ad Hoc Networks*, vol. 98, Mar. 2020, p. 102047. DOI.org (Crossref). <https://doi.org/10.1016/j.adhoc.2019.102047>.
- Alsina-Pagès, M., R., Llonch, P., Ginovart-Panisello, J., G., Guevara, R., Freixes, M., Castro, M., , Mainau, E. (2021). (2021). Dairy Cattle Welfare through Acoustic Analysis: Preliminary results of acoustic environment description. *Proceedings of Euronoise 2021*, 25–27 October, Madeira, Portugal..
- Arablouei, Reza, al., e. (2024). "Cattle Behavior Recognition from Accelerometer Data: Leveraging in-Situ Cross-Device Model Learning." *Computers and Electronics in Agriculture*, vol. 227, Dec. 2024, p. 109546. <https://doi.org/10.1016/j.compag.2024.109546>.
- Araújo, M., V., al., e. (2025). "AI-Powered Cow Detection in Complex Farm Environments." *Smart Agricultural Technology*, vol. 10, Mar. 2025, p. 100770. <https://doi.org/10.1016/j.atech.2025.100770>.
- Aubé, L., al., e. (2025). "Validation of Qualitative Behaviour Assessment for Dairy Cows at Pasture." *Applied Animal Behaviour Science*, vol. 283, Feb. 2025, p. 106489. <https://doi.org/10.1016/j.applanim.2024.106489>.
- Avanzato, R., Avondo, M., Beritelli, F., Franco, D., F., , Tumino, S. (2023). (2023). Detecting the Number of Bite Prehension of Grazing Cows in an Extensive System Using an Audio Recording Method. In *Proceedings of the 8th International Conference of Yearly Reports on Informatics, Mathematics, and Engineering (ICYRIME 2023)* (pp. 27–31). Naples, Italy. CEUR-WS.org, Vol. 3684..
- Bendel, Oliver, , Zbinden., N. (2024). "The Animal Whisperer Project." *Proceedings of the International Conference on Animal-Computer Interaction*, ACM, 2024, pp. 1–9. <https://doi.org/10.1145/3702336.3702347>.
- Bertelsen, Maja, , Jensen., M. B. (2023). "Comparing Weaning Methods in Dairy Calves with Different Dam Contact Levels." *Journal of Dairy Science*, vol. 106, no. 12, Dec. 2023, pp. 9598–612. <https://doi.org/10.3168/jds.2023-23393>.

- Bloch, Victor, al., e. (2023). "Development and Analysis of a CNN- and Transfer-Learning-Based Classification Model for Automated Dairy Cow Feeding Behavior Recognition from Accelerometer Data." *Sensors*, vol. 23, no. 5, Feb. 2023, p. 2611. <https://doi.org/10.3390/s23052611>.
- Brady, Beth, al., e. (2022). "Manatee Calf Call Contour and Acoustic Structure Varies by Species and Body Size." *Scientific Reports*, vol. 12, no. 1, Nov. 2022, p. 19597. <https://doi.org/10.1038/s41598-022-23321-7>.
- Burnham, Rianna. (2023). "Animal Calling Behaviours and What This Can Tell Us about the Effects of Changing Soundscapes." *Acoustics*, vol. 5, no. 3, July 2023, pp. 631–52. DOI.org (Crossref). <https://doi.org/10.3390/acoustics5030039>.
- Castillejo, Pedro, al., e. (2019). "The AFarCloud ECSEL Project." 2019 22nd Euromicro Conference on Digital System Design (DSD), IEEE, 2019, pp. 414–19. DOI.org (Crossref). <https://doi.org/10.1109/DSD.2019.00066>.
- Cetintav, Bekir, al., e. (2025). "Generative AI Meets Animal Welfare: Evaluating GPT-4 for Pet Emotion Detection." *Animals*, vol. 15, no. 4, Feb. 2025, p. 492. <https://doi.org/10.3390/ani15040492>.
- Chelotti, O., J., Martinez-Rau, L. S., al., e. (2024). "Livestock Feeding Behaviour: A Review on Automated Systems for Ruminant Monitoring." *Biosystems Engineering*, vol. 246, Oct. 2024, pp. 150–77. <https://doi.org/10.1016/j.biosystemseng.2024.08.003>.
- Chelotti, O., J., Vanrell, S. R., Martinez-Rau, L. S., al., e. (2023). "Using Segment-Based Features of Jaw Movements to Recognise Foraging Activities in Grazing Cattle." *Biosystems Engineering*, vol. 229, May 2023, pp. 69–84. <https://doi.org/10.1016/j.biosystemseng.2023.03.014>.
- Chelotti, O., J., Vanrell, S. R., Rau, L. S. M., al., e. (2020). "An Online Method for Estimating Grazing and Rumination Bouts Using Acoustic Signals in Grazing Cattle." *Computers and Electronics in Agriculture*, vol. 173, June 2020, p. 105443. <https://doi.org/10.1016/j.compag.2020.105443>.
- Chen, Yiming, arXiv, e. a. V. B. L. V. A., 2024, D., (2024). . <https://doi.org/10.48550/arXiv.2410.17196>.
- Clarke, A., J. (2024). "Bison Mother–Offspring Acoustic Communication." *Journal of Mammalogy*, edited by Timothy Smyser, vol. 105, no. 5, Sept. 2024, pp. 1182–89. <https://doi.org/10.1093/jmammal/gyae076>.
- Cornips, Leonie. (2024). "The Semiotic Repertoire of Dairy Cows." *Language in Society*, Oct. 2024, pp. 1–25. <https://doi.org/10.1017/S0047404524000460>.
- Dimov, Dimo, al., e. (2023). "Importance of Noise Hygiene in Dairy Cattle Farming—A Review." *Acoustics*, vol. 5, no. 4, Nov. 2023, pp. 1036–45. <https://doi.org/10.3390/acoustics5040059>.
- Dixhoorn, V., E., I. D., al., e. (2023). "Behavioral Patterns as Indicators of Resilience after Parturition in Dairy Cows." *Journal of Dairy Science*, vol. 106, no. 9, Sept. 2023, pp. 6444–63. <https://doi.org/10.3168/jds.2022-22891>.
- Eckhardt, Regina, al., e. (2024). "Modelling Climate Change Impacts on Cattle Behavior Using Generative Artificial Intelligence: A Pathway to Adaptive Live-stock Management." 2024 Anaheim, California July 28–31, 2024, American Society of Agricultural and Biological Engineers, 2024. <https://doi.org/10.13031/aim.202400377>.
- Eddicks, Matthias, al., e. (2024). "Monitoring of Respiratory Disease Patterns in a Multimicrobially Infected Pig Population Using Artificial Intelligence and Aggregate Samples." *Viruses*, vol. 16, no. 10, Oct. 2024, p. 1575. DOI.org (Crossref). <https://doi.org/10.3390/v16101575>.
- Eriksson, H., al., e. (2022). "Strategies for Keeping Dairy Cows and Calves Together – a Cross-Sectional Survey Study." *Animal*, vol. 16, no. 9, Sept. 2022, p. 100624. <https://doi.org/10.1016/j.animal.2022.100624>.
- Ferrero, Mariano, al., e. (2023). "A Full End-to-End Deep Approach for Detecting and Classifying Jaw Movements from Acoustic Signals in Grazing Cattle." *Engineering Applications of Artificial Intelligence*, vol. 121, May 2023, p. 106016. <https://doi.org/10.1016/j.engappai.2023.106016>.
- Fuchs, Patricia, al., e. (2024). "Stress Indicators in Dairy Cows Adapting to Virtual Fencing." *Journal of Animal Science*, vol. 102, Jan. 2024, p. skae024. <https://doi.org/10.1093/jas/skae024>.
- Gavojdian, Dinu, Lazebnik, T., arXiv, e. a. B. M. L. f. V. A. o. D. C. u. N. A. S., 2023, J., (2023). . <https://doi.org/10.48550/arXiv.2307.13994>.
- Gavojdian, Dinu, Mincu, M., al., e. (2024). "BovineTalk: Machine Learning for Vocalization Analysis of Dairy Cattle under the Negative Affective State of Isolation." *Frontiers in Veterinary Science*, vol. 11, Feb. 2024, p. 1357109. <https://doi.org/10.3389/fvets.2024.1357109>.

- Geng, Hongbo, al., e. (2024). "Motion Focus Global-Local Network: Combining Attention Mechanism with Micro Action Features for Cow Behavior Recognition." *Computers and Electronics in Agriculture*, vol. 226, Nov. 2024, p. 109399. <https://doi.org/10.1016/j.compag.2024.109399>.
- Gr, in, Temple. (2021). "The Visual, Auditory, and Physical Environment of Livestock Handling Facilities and Its Effect on Ease of Movement of Cattle, Pigs, and Sheep." *Frontiers in Animal Science*, vol. 2, Oct. 2021, p. 744207. <https://doi.org/10.3389/fanim.2021.744207>.
- Green, Alex, C., r., Clark, C. E. F., al., e. (2020). "Context-Related Variation in the Peripartum Vocalisations and Phonatory Behaviours of Holstein-Friesian Dairy Cows." *Applied Animal Behaviour Science*, vol. 231, Oct. 2020, p. 105089. <https://doi.org/10.1016/j.applanim.2020.105089>.
- Green, Alex, C., r., Lidfors, L. M., al., e. (2021). "Vocal Production in Postpartum Dairy Cows: Temporal Organization and Association with Maternal and Stress Behaviors." *Journal of Dairy Science*, vol. 104, no. 1, Jan. 2021, pp. 826–38. <https://doi.org/10.3168/jds.2020-18891>.
- Hagiwara, arXiv, M. A. A. V. E. B. o. S., 2022, O., (2022). . <https://doi.org/10.48550/arXiv.2210.14493>.
- Hagiwara, Masato, Cusimano, M., arXiv, e. a. M. A. V. t. S., 2022, O., (2022). . <https://doi.org/10.48550/arXiv.2210.10857>.
- Hagiwara, Masato, Hoffman, B., arXiv, e. a. B. T. B. o. A. S., 2022, O., (2022). . <https://doi.org/10.48550/arXiv.2210.12300>.
- Hasenpusch, P., al., e. (2024). "Dairy Cow Personality: Consistency in a Familiar Testing Environment." *JDS Communications*, vol. 5, no. 5, Sept. 2024, pp. 511–15. <https://doi.org/10.3168/jdsc.2023-0499>.
- Holinger, Mirjam, al., e. (2024). "Behavioural Changes to Moderate Heat Load in Grazing Dairy Cows under On-Farm Conditions." *Livestock Science*, vol. 279, Jan. 2024, p. 105376. <https://doi.org/10.1016/j.livsci.2023.105376>.
- Huang, Rongjie, Underst, e. a. A., ing, Speech, G., Music, Sound, , arXiv, T. H., 2023, A., (2023). . <https://doi.org/10.48550/arXiv.2304.12995>.
- Islam, Muhaiminul, M., , Scott, S. D. (2022). "Exploring the Effects of Precision Livestock Farming Notification Mechanisms on Canadian Dairy Farmers." *Science and Technologies for Smart Cities*, edited by Sara Paiva et al., vol. 442, Springer International Publishing, 2022, pp. 247–66. DOI.org (Crossref). [https://doi.org/10.1007/978-3-031-06371-8\\_16](https://doi.org/10.1007/978-3-031-06371-8_16).
- Jobarteh, Bubacarr, Acoustic, e. a. M. M. I. F. o., arXiv, L. D. f. D. D. C. V. i. A. W. A., DOI.org, 2. (2024). (Datacite). <https://doi.org/10.48550/ARXIV.2411.00477>.
- Johnsen, Føske, J., Johanssen, J. R. E., al., e. (2021). "Investigating Cow-calf Contact in Cow-Driven Systems: Behaviour of the Dairy Cow and Calf." *Journal of Dairy Research*, vol. 88, no. 1, Feb. 2021, pp. 52–55. <https://doi.org/10.1017/S0022029921000194>.
- Johnsen, Føske, J., Sørby, J., al., e. (2024). "Effect of Debonding on Stress Indicators in Cows and Calves in a Cow-Calf Contact System." *JDS Communica-tions*, vol. 5, no. 5, Sept. 2024, pp. 426–30. <https://doi.org/10.3168/jdsc.2023-0468>.
- Jung, Dae-Hyun, Kim, N. Y., Moon, S. H., Jhin, C., al., e. (2021). "Deep Learning-Based Cattle Vocal Classification Model and Real-Time Livestock Monitoring System with Noise Filtering." *Animals*, vol. 11, no. 2, Feb. 2021, p. 357. <https://doi.org/10.3390/ani11020357>.
- Jung, Dae-Hyun, Kim, N. Y., Moon, S. H., Kim, H. S., al., e. (2021). "Classification of Vocalization Recordings of Laying Hens and Cattle Using Convolutional Neural Network Models." *Journal of Biosystems Engineering*, vol. 46, no. 3, Sept. 2021, pp. 217–24. <https://doi.org/10.1007/s42853-021-00101-1>.
- Karmiris, Ilias, al., e. (2021). "Estimating Livestock Grazing Activity in Remote Areas Using Passive Acoustic Monitoring." *Information*, vol. 12, no. 8, July 2021, p. 290. <https://doi.org/10.3390/info12080290>.
- Kim, Eunbeen, al., e. (2023). "DualDiscWaveGAN-Based Data Augmentation Scheme for Animal Sound Classification." *Sensors*, vol. 23, no. 4, Feb. 2023, p. 2024. DOI.org (Crossref). <https://doi.org/10.3390/s23042024>.
- Kok, Akke, al., e. (2023). "Do You See the Pattern? Make the Most of Sensor Data in Dairy Cows." *Journal of Dairy Research*, vol. 90, no. 3, Aug. 2023, pp. 252–56. <https://doi.org/10.1017/S0022029923000559>.
- Lamanna, Martina, Bovo, M., , Applica-tions, D. C. ". C. T. f. D. C. A. S. R. o. t. C., 15.3, F. I. i. P. L. F. A. a. O. A. J. f. M. (2025). (2025): 458. <https://doi.org/10.3390/ani15030458>

- Lange, Annika, al., e. (2020). "Talking to Cows: Reactions to Different Auditory Stimuli During Gentle Human-Animal Interactions." *Frontiers in Psychology*, vol. 11, Oct. 2020, p. 579346. <https://doi.org/10.3389/fpsyg.2020.579346>.
- Lardy, R., al., e. (2022). "Understanding Anomalies in Animal Behaviour: Data on Cow Activity in Relation to Health and Welfare." *Animal - Open Space*, vol. 1, no. 1, Dec. 2022, p. 100004. <https://doi.org/10.1016/j.anopes.2022.100004>.
- Laurijs, A., K., al., e. (2021). "Vocalisations in Farm Animals: A Step towards Positive Welfare Assessment." *Applied Animal Behaviour Science*, vol. 236, Mar. 2021, p. 105264. <https://doi.org/10.1016/j.applanim.2021.105264>.
- Lefèvre, A., R., al., e. (2025). "Machine Learning Algorithms Can Predict Emotional Valence across Ungulate Vocalizations." *iScience*, vol. 28, no. 2, Feb. 2025, p. 111834. <https://doi.org/10.1016/j.isci.2025.111834>.
- Lenner, Ádám, al., e. (2023). "Calming Hungarian Grey Cattle in Headlocks Using Processed Nasal Vocalization of a Mother Cow." *Animals*, vol. 14, no. 1, Dec. 2023, p. 135. <https://doi.org/10.3390/ani14010135>.
- Lenner, Ádám, al., e. (2025). "Analysis of Sounds Made by Bos Taurus and Bubalus Bubalis Dams to Their Calves." *Frontiers in Veterinary Science*, vol. 12, Mar. 2025, p. 1549100. DOI.org (Crossref). <https://doi.org/10.3389/fvets.2025.1549100>.
- Li, Baihan, arXiv, e. a. D. L. A. T. C. f. D. A. G., 2024, J., (2024). . <https://doi.org/10.48550/arXiv.2407.13198>.
- Li, Chao, Minati, L., al., e. (2022). "Integrated Data Augmentation for Accelerometer Time Series in Behavior Recognition: Roles of Sampling, Balancing, and Fourier Surrogates." *IEEE Sensors Journal*, vol. 22, no. 24, Dec. 2022, pp. 24230–41. <https://doi.org/10.1109/JSEN.2022.3219594>.
- Li, Chao, Tokgoz, K., al., e. (2021). "Data Augmentation for Inertial Sensor Data in CNNs for Cattle Behavior Classification." *IEEE Sensors Letters*, vol. 5, no. 11, Nov. 2021, pp. 1–4. <https://doi.org/10.1109/LSENS.2021.3119056>.
- Li, Guoming, al., e. (2021). "Classifying Ingestive Behavior of Dairy Cows via Automatic Sound Recognition." *Sensors*, vol. 21, no. 15, Aug. 2021, p. 5231. <https://doi.org/10.3390/s21155231>.
- Linstädt, Jenny, al., e. (2024). "Animal-Based Welfare Indicators for Dairy Cows and Their Validity and Practicality: A Systematic Review of the Existing Literature." *Frontiers in Veterinary Science*, vol. 11, July 2024, p. 1429097. <https://doi.org/10.3389/fvets.2024.1429097>.
- Liu, Jiefei, al., e. (2024). "Development of a Novel Classification Approach for Cow Behavior Analysis Using Tracking Data and Unsupervised Machine Learning Techniques." *Sensors*, vol. 24, no. 13, June 2024, p. 4067. <https://doi.org/10.3390/s24134067>.
- Liu, ting, T., al., e. (2020). "Development Process of Animal Image Recognition Technology and Its Application in Modern Cow and Pig Industry." *IOP Conference Series: Earth and Environmental Science*, vol. 512, no. 1, June 2020, p. 012090. <https://doi.org/10.1088/1755-1315/512/1/012090>.
- Mac, E., S., al., e. (2023). "Behavioral Responses to Cow and Calf Separation: Separation at 1 and 100 Days after Birth." *Animal Bioscience*, vol. 36, no. 5, May 2023, pp. 810–17. <https://doi.org/10.5713/ab.22.0257>.
- Mahmud, Sultan, M., al., e. (2021). "A Systematic Literature Review on Deep Learning Applications for Precision Cattle Farming." *Computers and Electronics in Agriculture*, vol. 187, Aug. 2021, p. 106313. <https://doi.org/10.1016/j.compag.2021.106313>.
- Martinez-Rau, S., L., Chelotti, J. O., Ferrero, M., Galli, J. R., al., e. (2025). "A Noise-Robust Acoustic Method for Recognizing Foraging Activities of Grazing Cattle." *Computers and Electronics in Agriculture*, vol. 229, Feb. 2025, p. 109692. <https://doi.org/10.1016/j.compag.2024.109692>.
- Martinez-Rau, S., L., Chelotti, J. O., Ferrero, M., Utsumi, S. A., al., e. (2023). "Daylong Acoustic Recordings of Grazing and Rumination Activities in Dairy Cows." *Scientific Data*, vol. 10, no. 1, Nov. 2023, p. 782. <https://doi.org/10.1038/s41597-023-02673-3>.
- Martinez-Rau, S., L., Chelotti, J. O., Giovanini, L. L., al., e. (2024). "On-Device Feeding Behavior Analysis of Grazing Cattle." *IEEE Transactions on Instrumentation and Measurement*, vol. 73, 2024, pp. 1–13. <https://doi.org/10.1109/TIM.2024.3376013>.
- Martinez-Rau, S., L., Chelotti, J. O., Vanrell, S. R., al., e. (2022). "A Robust Computational Approach for Jaw Movement Detection and Classification in Grazing Cattle Using Acoustic Signals." *Computers and Electronics in Agriculture*, vol. 192, Jan. 2022, p. 106569. <https://doi.org/10.1016/j.compag.2021.106569>.

- Martinez-Rau, Sebastian, L., al., e. (2023). "Real-Time Acoustic Monitoring of Foraging Behavior of Grazing Cattle Using Low-Power Embedded Devices." 2023 IEEE Sensors Applications Symposium (SAS), IEEE, 2023, pp. 01–06. <https://doi.org/10.1109/SAS58821.2023.10254175>.
- Marumo, L., J., al., e. (2024). "Behavioural Variability, Physical Activity, Rumination Time, and Milk Characteristics of Dairy Cattle in Response to Regrouping." *Animal*, vol. 18, no. 3, Mar. 2024, p. 101094. <https://doi.org/10.1016/j.animal.2024.101094>.
- McManus, Rosemary, al., e. (2022). "Thermography for Disease Detection in Livestock: A Scoping Review." *Frontiers in Veterinary Science*, vol. 9, Aug. 2022, p. 965622. DOI.org (Crossref). <https://doi.org/10.3389/fvets.2022.965622>.
- Meen, H., G., al., e. (2015). "Sound Analysis in Dairy Cattle Vocalisation as a Potential Welfare Monitor." *Computers and Electronics in Agriculture*, vol. 118, Oct. 2015, pp. 111–15. DOI.org. <https://doi.org/10.1016/j.compag.2015.08.028>.
- Mehdizadeh, A., Saman, al., e. (2023). "Classifying Chewing and Rumination in Dairy Cows Using Sound Signals and Machine Learning." *Animals*, vol. 13, no. 18, Sept. 2023, p. 2874. <https://doi.org/10.3390/ani13182874>.
- Miao, Zhongqi, Review, e. a. Z. T. f. W. B. D. I., 2023, A., (2023). . <https://doi.org/10.21203/rs.3.rs-3180218/v1>.
- Miron, Marius, arXiv, e. a. B. A. V. D. w. A. t. C. D., 2025, M., (2025). . <https://doi.org/10.48550/arXiv.2410.03427>.
- Moreira, Madruga, S., al., e. (2023). "Auditory Sensitivity in Beef Cattle of Different Genetic Origin." *Journal of Veterinary Behavior*, vol. 59, Jan. 2023, pp. 67–72. DOI.org (Crossref). <https://doi.org/10.1016/j.jveb.2022.10.004>.
- Moutaouakil, E., Khalid, , Falih, N. (2023). "A Design of a Smart Farm System for Cattle Monitoring." *Indonesian Journal of Electrical Engineering and Computer Science*, vol. 32, no. 2, Nov. 2023, p. 857. DOI.org (Crossref). <https://doi.org/10.11591/ijeecs.v32.i2.pp857-864>.
- Neethirajan, 6.4, S. ". a. L. T. M. t. D. C. V. A. N. A. A. t. P. W. A. (2025). (2025): 65. <https://doi.org/10.3390/ai6040065>
- Neethirajan, Suresh. (2022). "Affective State Recognition in Livestock—Artificial Intelligence Approaches." *Animals*, vol. 12, no. 6, Mar. 2022, p. 759. <https://doi.org/10.3390/ani12060759>.
- Noda, T., Koizumi, T., Yukitake, sources, N. e. a. A. s. l. a. a. s. f. e. c. o. s., 14, d. t. f. m. n. u. s. S. R., 6394 (6394). (2024). <https://doi.org/10.1038/s41598-024-56439-x>
- Nolasco, Ines, al., e. (2023). "Learning to Detect an Animal Sound from Five Examples." *Ecological Informatics*, vol. 77, Nov. 2023, p. 102258. <https://doi.org/10.1016/j.ecoinf.2023.102258>.
- Ntalampiras, Stavros, al., e. (2020). "Automatic Detection of Cow/Calf Vocalizations in Free-Stall Barn." 2020 43rd International Conference on Telecommunications and Signal Processing (TSP), IEEE, 2020, pp. 41–45. DOI.org. <https://doi.org/10.1109/TSP49548.2020.9163522>.
- Nurcholis, , Sumaryanti., L. (2021). "Reproductive Behavior's: Audiovisual Detection of Oestrus after Synchronization Using Prostaglandin F2 Alpha (PGF2α)." *E3S Web of Conferences*, edited by I.H.A. Wahab et al., vol. 328, 2021, p. 04021. <https://doi.org/10.1051/e3sconf/202132804021>.
- Oestreich, K., W., al., e. (2024). "Listening to Animal Behavior to Understand Changing Ecosystems." *Trends in Ecology & Evolution*, vol. 39, no. 10, Oct. 2024, pp. 961–73. <https://doi.org/10.1016/j.tree.2024.06.007>.
- Özmen, Güzin, al., e. (2022). "Sound Analysis to Recognize Cattle Vocalization in a Semi-Open Barn." *Gazi Journal of Engineering Sciences*, vol. 8, no. 1, May 2022, pp. 158–67. <https://doi.org/10.30855/gmbd.2022.01.15>.
- P, eya, Raj, Y., Bhattarai, B., , Lee., J. (2020). "Visual Object Detector for Cow Sound Event Detection." *IEEE Access*, vol. 8, 2020, pp. 162625–33. <https://doi.org/10.1109/ACCESS.2020.3022058>.
- P, eya, Raj, Y., Bhattarai, B., Afzaal, U., al., e. (2022). "A Monophonic Cow Sound Annotation Tool Using a Semi-Automatic Method on Audio/Video Data." *Livestock Science*, vol. 256, Feb. 2022, p. 104811. <https://doi.org/10.1016/j.livsci.2021.104811>.
- Page, J., M., al., e. (2020). "PRISMA 2020 Explanation and Elaboration: Updated Guidance and Exemplars for Reporting Systematic Reviews." *BMJ*, Mar. 2021, p. n160. DOI.org (Crossref). <https://doi.org/10.1136/bmj.n160>.

- Pallottino, Federico, al., e. (2025). "Applications and Perspectives of Generative Artificial Intelligence in Agriculture." *Computers and Electronics in Agriculture*, vol. 230, Mar. 2025, p. 109919. <https://doi.org/10.1016/j.compag.2025.109919>.
- Patil, Raturaj, al., e. (2024). "Identifying Indian Cattle Behaviour Using Acoustic Biomarkers:" *Proceedings of the 13th International Conference on Pattern Recognition Applications and Methods, SCITEPRESS - Science and Technology Publications, 2024*, pp. 594–602. <https://doi.org/10.5220/0012576800003654>.
- Pei, Te, Clustering, e. a. G. A. D. T. F. I. H., arXiv, L. L. M. f. E. C., DOI.org, 2. (2025). (Datacite). <https://doi.org/10.48550/ARXIV.2501.13743>.
- Peng, Yingqi, Chen, Y., al., e. (2024). "A Multimodal Classification Method: Cow Behavior Pattern Classification with Improved EdgeNeXt Using an Inertial Measurement Unit." *Computers and Electronics in Agriculture*, vol. 226, Nov. 2024, p. 109453. <https://doi.org/10.1016/j.compag.2024.109453>.
- Peng, Yingqi, Wul, ari, al., e. (2023). "Japanese Black Cattle Call Patterns Classification Using Multiple Acoustic Features and Machine Learning Models." *Computers and Electronics in Agriculture*, vol. 204, Jan. 2023, p. 107568. <https://doi.org/10.1016/j.compag.2022.107568>.
- Prestegaard-Wilson, Jacquelyn, , Vitale., J. (2024). "Generative Artificial Intelligence in Extension: A New Era of Support for Livestock Producers." *Animal Frontiers*, vol. 14, no. 6, Dec. 2024, pp. 57–59. <https://doi.org/10.1093/af/vfae024>.
- Pérez-Granados, Cristian, , Schuchmann., K. (2023). "The Sound of the Illegal: Applying Bioacoustics for Long-Term Monitoring of Illegal Cattle in Protected Areas." *Ecological Informatics*, vol. 74, May 2023, p. 101981. <https://doi.org/10.1016/j.ecoinf.2023.101981>.
- Pérez-Torres, L., al., e. (2021). "Short- and Long-Term Effects of Temporary Early Cow–Calf Separation or Restricted Suckling on Well-Being and Performance in Zebu Cattle." *Animal*, vol. 15, no. 2, Feb. 2021, p. 100132. <https://doi.org/10.1016/j.animal.2020.100132>.
- Radford, Alec, 2022, e. a. R. S. R. v. L. W. S., <https://arxiv.org/abs/2212.04356>. (2022). .
- Ramos, Angel, E., al., e. (2023). "Antillean Manatee Calves in Captive Rehabilitation Change Vocal Behavior in Anticipation of Feeding." *Zoo Biology*, vol. 42, no. 6, Nov. 2023, pp. 723–29. <https://doi.org/10.1002/zoo.21785>.
- Riaboff, L., al., e. (2022). "Predicting Livestock Behaviour Using Accelerometers: A Systematic Review of Processing Techniques for Ruminant Behaviour Prediction from Raw Accelerometer Data." *Computers and Electronics in Agriculture*, vol. 192, Jan. 2022, p. 106610. <https://doi.org/10.1016/j.compag.2021.106610>.
- Robinson, David, arXiv, e. a. N. A. A. F. M. f. B., 2024, N., (2024). . <https://doi.org/10.48550/arXiv.2411.07186>.
- Rohan, Ali, al., e. (2024). "Application of Deep Learning for Livestock Behaviour Recognition: A Systematic Literature Review." *Computers and Electronics in Agriculture*, vol. 224, Sept. 2024, p. 109115. <https://doi.org/10.1016/j.compag.2024.109115>.
- Rubenstein, K., P., Speak, e. a. A. A. L. L. M. T. C., arXiv, L., 2023, J., (2023). . <https://doi.org/10.48550/arXiv.2306.12925>.
- Russel, Shebiah, N., , Selvaraj., A. (2024). "Decoding Cow Behavior Patterns from Accelerometer Data Using Deep Learning." *Journal of Veterinary Behavior*, vol. 74, July 2024, pp. 68–78. <https://doi.org/10.1016/j.jvbeh.2024.06.005>.
- Röttgen, V., al., e. (2020). "Automatic Recording of Individual Oestrus Vocalisation in Group-Housed Dairy Cattle: Development of a Cattle Call Monitor." *Animal*, vol. 14, no. 1, 2020, pp. 198–205. <https://doi.org/10.1017/S1751731119001733>.
- Sattar, Farook. (2022). "A Context-Aware Method-Based Cattle Vocal Classification for Livestock Monitoring in Smart Farm." *The 1st International Online Conference on Agriculture&mdash;Advances in Agricultural Science and Technology, MDPI, 2022*, p. 89. <https://doi.org/10.3390/IOCAG2022-12233>.
- Schillings, J., al., e. (2024). "Managing End-User Participation for the Adoption of Digital Livestock Technologies: Expectations, Performance, Relationships, and Support." *The Journal of Agricultural Education and Extension*, vol. 30, no. 2, Mar. 2024, pp. 277–95. DOI.org (Crossref). <https://doi.org/10.1080/1389224X.2023.2192697>.

- Schnaider, Alice, M., al., e. (2022). "Vocalization and Other Behaviors Indicating Pain in Beef Calves during the Ear Tagging Procedure." *Journal of Veterinary Behavior*, vol. 47, Jan. 2022, pp. 93–98. <https://doi.org/10.1016/j.jveb.2021.10.005>.
- Schnaider, Ma, al., e. (2022). "Vocalization and Other Behaviors as Indicators of Emotional Valence: The Case of Cow-Calf Separation and Reunion in Beef Cattle." *Journal of Veterinary Behavior*, vol. 49, Mar. 2022, pp. 28–35. <https://doi.org/10.1016/j.jveb.2021.11.011>.
- Sert, P. d., N., Hurst, V., Ahluwalia, 16, A. e. a. T. A. g. 2. U. g. f. r. a. r. B. V. R., 242 (2020). (2020). <https://doi.org/10.1186/s12917-020-02451-y>
- Seyfarth, M., R., , Cheney., D. L. (2003). "Signalers and Receivers in Animal Communication." *Annual Review of Psychology*, vol. 54, no. 1, Feb. 2003, pp. 145–73. DOI.org. <https://doi.org/10.1146/annurev.psych.54.101601.145121>.
- Sharma, Sug, ha, , Kadyan., V. (2023). "Detection of Estrus through Automated Classification Approaches Using Vocalization Pattern in Murrah Buffa-loes." *2023 3rd International Conference on Artificial Intelligence and Signal Processing (AISP)*, IEEE, 2023, pp. 1–6. <https://doi.org/10.1109/AISP57993.2023.10134787>.
- Shi, Zhonghao, al., e. (2024). "Classifying and Understanding of Dairy Cattle Health Using Wearable Inertial Sensors With Random Forest and Explainable Artificial Intelligence." *IEEE Sensors Letters*, vol. 8, no. 3, Mar. 2024, pp. 1–4. <https://doi.org/10.1109/LENS.2024.3358619>.
- Shorten, R., P. (2023). "Acoustic Sensors for Detecting Cow Behaviour." *Smart Agricultural Technology*, vol. 3, Feb. 2023, p. 100071. <https://doi.org/10.1016/j.atech.2022.100071>.
- Shorten, R., P., , Hunter., L. B. (2023). "Acoustic Sensors for Automated Detection of Cow Vocalization Duration and Type." *Computers and Electronics in Agriculture*, vol. 208, May 2023, p. 107760. <https://doi.org/10.1016/j.compag.2023.107760>.
- Shorten, R., P., , Hunter., L. B. (2024). "Acoustic Sensors to Detect the Rate of Cow Vocalization in a Complex Farm Environment." *Applied Animal Behaviour Science*, vol. 278, Sept. 2024, p. 106377. <https://doi.org/10.1016/j.applanim.2024.106377>.
- Silva, D., Martins, M., al., e. (2024). "Acoustic-Based Models to Assess Herd-Level Calves' Emotional State: A Machine Learning Approach." *Smart Agricultural Technology*, vol. 9, Dec. 2024, p. 100682. <https://doi.org/10.1016/j.atech.2024.100682>.
- Slob, Naftali, al., e. (2021). "Application of Machine Learning to Improve Dairy Farm Management: A Systematic Literature Review." *Preventive Veterinary Medicine*, vol. 187, Feb. 2021, p. 105237. <https://doi.org/10.1016/j.prevetmed.2020.105237>.
- Stachowicz, Joanna, al., e. (2022). "Can We Detect Patterns in Behavioral Time Series of Cows Using Cluster Analysis?" *Journal of Dairy Science*, vol. 105, no. 12, Dec. 2022, pp. 9971–81. <https://doi.org/10.3168/jds.2022-22140>.
- Stowell, Dan. (2022). "Computational Bioacoustics with Deep Learning: A Review and Roadmap." *PeerJ*, vol. 10, Mar. 2022, p. e13152. <https://doi.org/10.7717/peerj.13152>.
- Stygar, H., A., al., e. (2022). "How Far Are We From Data-Driven and Animal-Based Welfare Assessment? A Critical Analysis of European Quality Schemes." *Frontiers in Animal Science*, vol. 3, May 2022, p. 874260.. <https://doi.org/10.3389/fanim.2022.874260>.
- Sun, Yifei, al., e. (2023). "Free-Ranging Livestock Changes the Acoustic Properties of Summer Soundscapes in a Northeast Asian Temperate Forest." *Biological Conservation*, vol. 283, July 2023, p. 110123. <https://doi.org/10.1016/j.biocon.2023.110123>.
- Takefuji, Yoshiyasu. (2024). "Unveiling Livestock Trade Trends: A Beginner's Guide to Generative AI-Powered Visualization." *Research in Veterinary Science*, vol. 180, Nov. 2024, p. 105435. <https://doi.org/10.1016/j.rvsc.2024.105435>.
- Torre, P. D. L., Mónica, al., e. (2015). "Acoustic Analysis of Cattle (*Bos Taurus*) Mother–Offspring Contact Calls from a Source–Filter Theory Perspective." *Applied Animal Behaviour Science*, vol. 163, Feb. 2015, pp. 58–68. DOI.org. <https://doi.org/10.1016/j.applanim.2014.11.017>.
- Tuytens, AM, F., Molento, C. F., , farming, S. B. ". t. o. p. l. (2022). (PLF) for animal welfare." *Frontiers in Veterinary Science* 9 (2022): 889623. <https://doi.org/10.3389/fvets.2022.889623>

- V, ermeulen, Joris, al., e. (2016). "Early Recognition of Bovine Respiratory Disease in Calves Using Automated Continuous Monitoring of Cough Sounds." *Computers and Electronics in Agriculture*, vol. 129, Nov. 2016, pp. 15–26. DOI.org (Crossref). <https://doi.org/10.1016/j.compag.2016.07.014>.
- Vidal, Gema, al., e. (2023). "Comparative Performance Analysis of Three Machine Learning Algorithms Applied to Sensor Data Registered by a Leg-Attached Accelerometer to Predict Metritis Events in Dairy Cattle." *Frontiers in Animal Science*, vol. 4, Apr. 2023, p. 1157090. <https://doi.org/10.3389/fanim.2023.1157090>.
- Vidaña-Vila, E., Malé, J., Freixes, M., Solís-Cifré, M., Jiménez, M., Larrondo, C., Guevara, R., Mir, a, J., Duboc, L., Mainau, E., Llonch, P., , Alsina-Pagès, M., R. (2023). (2023). Automatic Detection of Cow Vocalizations Using Convolutional Neural Networks. In *Proceedings of the 8th Detection and Classification of Acoustic Scenes and Events Workshop (DCASE2023)* (pp. 206–210). Tampere, Finland..
- Vogt, Anina, al., e. (2025). "Dairy Cows' Responses to 2 Separation Methods after 3 Months of Cow-Calf Contact." *Journal of Dairy Science*, vol. 108, no. 2, Feb. 2025, pp. 1940–63. <https://doi.org/10.3168/jds.2024-25293>.
- Volkman, N., al., e. (2021). "On-Farm Detection of Claw Lesions in Dairy Cows Based on Acoustic Analyses and Machine Learning." *Journal of Dairy Science*, vol. 104, no. 5, May 2021, pp. 5921–31. <https://doi.org/10.3168/jds.2020-19206>.
- Vranken, E., Mounir, M., Norton, T. (2023). (2023). Sound-Based Monitoring of Livestock. In: Zhang, Q. (eds) *Encyclopedia of Digital Agricultural Technologies*. Springer, Cham. [https://doi.org/10.1007/978-3-031-24861-0\\_129](https://doi.org/10.1007/978-3-031-24861-0_129)
- Vu, H., Prabhune, O., Raskar, U., P, itharatne, D., Chung, H., Choi, Y., C., , Kim, Y. (2024). (2024). MmCows: A Multimodal Dataset for Dairy Cattle Monitoring. In *Advances in Neural Information Processing Systems*, 37, 59451–59467..
- Wang, Bin, arXiv, e. a. A. A. U. B. f. A. L. L. M., 2024, N., (2024). . <https://doi.org/10.48550/arXiv.2406.16020>.
- Wang, Jun, Chen, H., al., e. (2023). "Identification of Oestrus Cows Based on Vocalisation Characteristics and Machine Learning Technique Using a Du-al-Channel-Equipped Acoustic Tag." *Animal*, vol. 17, no. 6, June 2023, p. 100811. <https://doi.org/10.1016/j.animal.2023.100811>.
- Wang, Jun, Si, Y., al., e. (2023). "Discrimination Strategy Using Machine Learning Technique for Oestrus Detection in Dairy Cows by a Dual-Channel-Based Acoustic Tag." *Computers and Electronics in Agriculture*, vol. 210, July 2023, p. 107949. <https://doi.org/10.1016/j.compag.2023.107949>.
- Wang, Ziwei, al., e. (2022). "Multi-Modal Sensing for Behaviour Recognition." *Proceedings of the 28th Annual International Conference on Mobile Computing And Networking*, ACM, 2022, pp. 900–02. <https://doi.org/10.1145/3495243.3558277>.
- Watts, M., J., , Stookey., J. M. (2000). "Vocal Behaviour in Cattle: The Animal's Commentary on Its Biological Processes and Welfare." *Applied Animal Behaviour Science*, vol. 67, no. 1–2, Mar. 2000, pp. 15–33. DOI.org. [https://doi.org/10.1016/S0168-1591\(99\)00108-2](https://doi.org/10.1016/S0168-1591(99)00108-2).
- Welk, Allison, al., e. (2024). "Invited Review: The Effect of Weaning Practices on Dairy Calf Performance, Behavior, and Health—A Systematic Review." *Journal of Dairy Science*, vol. 107, no. 8, Aug. 2024, pp. 5237–58. <https://doi.org/10.3168/jds.2024-24521>.
- Wu, Haibin, arXiv, e. a. T. A. L. M. –. a. O., 2024, F., (2024). . <https://doi.org/10.48550/arXiv.2402.13236>.
- Wu, Yiqi, Underst, e. a. G. V. P. P. o. M. L. L. M. i. P. A., arXiv, i., 2024, J., (2024). . <https://doi.org/10.48550/arXiv.2406.09781>.
- Yang, Qian, arXiv, e. a. A. B. L. A. M. v. G. C., 2024, J., (2024). . <https://doi.org/10.48550/arXiv.2402.07729>.
- Yoshihara, Yu, , Oya., K. (2021). "Characterization and Assessment of Vocalization Responses of Cows to Different Physiological States." *Journal of Applied Animal Research*, vol. 49, no. 1, Jan. 2021, pp. 347–51. <https://doi.org/10.1080/09712119.2021.1967756>.
- Z, Baig, T., , Ch, Shastry., r. (2022). "Parturition and Estrus Detection in Cows and Heifers with WSN and IoT." *2022 2nd International Conference on Technological Advancements in Computational Sciences (ICTACS)*, IEEE, 2022, pp. 201–08. <https://doi.org/10.1109/ICTACS56270.2022.9987783>.

**Disclaimer/Publisher's Note:** The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s)

disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.