Preprints.org

Review

AI-Driven Bioacoustics in Poultry Farming: A Critical Systematic Review on Vocalization Analysis for Stress and Disease Detection

Venkatraman Manikandan and Suresh Neethirajan [*]

Review

AI-Driven Bioacoustics in Poultry Farming: A Critical Systematic Review on Vocalization Analysis for Stress and Disease Detection

Venkatraman Manikandan [1] and Suresh Neethirajan [1,2,*]

[1] Faculty of Computer Science, Dalhousie University, Halifax, NS B3H 4R2, Canada; venkatraman.m@dal.ca

[2] Faculty of Agriculture, Dalhousie University, Halifax, NS B3H 4R2, Canada

* Correspondence: sneethir@gmail.com

Abstract: The fusion of artificial intelligence (AI) and acoustic sensing is fundamentally reshaping poultry welfare monitoring by offering unprecedented, non-invasive insights into the emotional, physiological, and behavioral states of birds through vocal analysis. This systematic review delves into the transformative intersection of bioacoustics, machine learning, and animal welfare, meticulously examining the shift from traditional acoustic feature extraction methods, like Mel-Frequency Cepstral Coefficients (MFCCs) and spectrogram analysis, toward cutting-edge deep learning architectures, such as CNNs, LSTMs, attention mechanisms, and powerful self-supervised models like wav2vec2 and Whisper. Critically, it highlights the evolution toward compact, real-time deployment solutions, including TinyML and edge computing, designed specifically for the dynamic and noisy environments of commercial poultry farms. Emotion recognition, early disease detection, and nuanced behavioral decoding emerge as pivotal application areas, underscoring the immense potential for proactive and responsive livestock management. Furthermore, the review sheds light on emerging open-source toolkits and automated pipelines enhancing dataset preprocessing, annotation accuracy, and robust acoustic inference. Employing advanced bibliometric mapping and thematic clustering, the paper identifies critical gaps in research reproducibility, dataset standardization, and the interpretability of complex AI models. It calls for rigorous advancements in explainable AI methodologies, advocates for greater cross-species generalization of acoustic models, and urges a conscientious approach to ethical AI design. Ultimately, this review underscores the necessity for holistic, interpretable, and ethically informed frameworks, positioning acoustic AI as a transformative force in enhancing animal welfare and operational efficiency within the poultry industry.

Keywords: poultry vocalization; acoustic monitoring; edge AI; TinyML; animal welfare; bioacoustics classification

1. Introduction

The integration of artificial intelligence (AI) into livestock monitoring is reshaping the landscape of animal welfare, behavior analysis, and environmental control. Acoustic sensing is one of several potential modalities and has emerged as a powerful, non-invasive way of monitoring the physiological and emotional states of animals-in, poultry. Vocalizations carry a wealth of biological and behavioral information, and, if properly captured, preprocessed, and analyzed, they serve as digital biomarkers with respect to stress, disease, environmental discomfort, and even social or emotional cues. This systematic study in the literature explores the convergence of bioacoustics, machine learning (ML), and animal welfare with respect to poultry vocalizations as the primary data modality. While resourceful and providing a foundation for exploration, traditional methods, especially Mel Frequency Cepstral Coefficients (MFCC) and spectrogram analysis, are being replaced by rapid advances in deep learning, transfer learning, and self-supervised audio models.

Additionally, TinyML, edge computing, and real-time deployment frameworks have brought these models closer to practical farm-level applications. Chickens have more than 30 different types of calls that span from distress, mating, predator threats etc. This makes their vocal repertoires one of the most diverse kinds among domesticated animals. These repertoires may give insights into their emotional and physiological states, thus making vocalization analysis one of the most powerful and non-invasive methods to identify their welfare and state. Vocalizations, from the ethological and communication theory viewpoint, tend to be the selected evolutionary tools for social coordination developed by environmental pressures and flock dynamics. Analyzing poultry vocalizations in that sense aligns with embodied cognition, whereby vocal behavior extends beyond just signaling but becomes a reflection of internal state and context. This review identifies the trend in methodologies used and key benchmark architectures through a comprehensive thematic synthesis of peer-reviewed studies and identifies critical gaps in current approaches. Increasing importance is put on multi-modal and explainable AI; the dynamic acoustic features rather than static; and standardized datasets and pipelines for reproducibility and generalization. Furthermore, this work adds bibliometric co-occurrence mapping to illustrate evolving thematic structure in the field, thereby aiding in identifying future research trajectories and interdisciplinary collaborations. By bridging computational modeling with ethological relevance, this review aims to inform researchers, practitioners, and technologists about the current state, limitations, and untapped potential of AI-driven poultry vocalization analysis. The review entails a systematic search approach through IEEE Xplore, PubMed, Scopus, Web of Science, SpringerLink, etc., focusing on research work done between 2018 and March 2025. The query consisted of various terms related to poultry vocalizations and AI (e.g., "chicken," "acoustic," "machine learning," "CNN," "Transformer," "wav2vec").
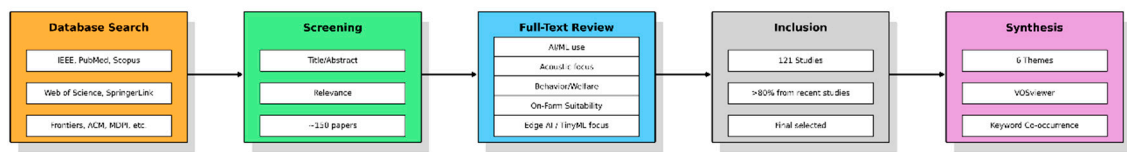

Figure 1. Systematic review pipeline outlining database search, screening, full-text evaluation for on-farm AI acoustic studies, and thematic synthesis from 121 included papers.

A total of approximately 150 research works were analyzed, and out of these, 121 were judged to be relevant with respect to technical rigor and contribution toward poultry acoustic sensing. Studies using machine learning or signal processing applied to vocalizations concerning behavior, welfare, or disease monitoring were included in this review. Seminal references related to acoustic features and deep learning (e.g., MFCCs, attention models) were retained as background for the technical context. The complementary papers are classified into six main themes: acoustic features, ML/DL models, behavior and stress detection, disease classification, toolkits and pipelines, and on-farm deployment. More than 85% of the references were published between 2020 and 2025, illustrating the fast-paced growth of this multidisciplinary field.

2. Acoustic Features and Preprocessing Techniques

The meaningful extraction of acoustic features and sound preprocessing techniques are pivotal in animal vocalization analysis. All the reviewed literature indicates that MFCCs, STFT, spectral entropy, and Mel-spectrograms have always been the core components of both traditional and deep learning pipelines. The most popular acoustic feature is the MFCC, which has been cited in over half

of the papers for the classification of animal sounds. They have been used to characterize vocational sounds from broiler birds, laying hens, chicks, and ducks and the other species as perceptually relevant frequency information is extracted. For example, Umarani et al. [1], Pereira et al. [2], Jung et al. [3], and Thomas et al. [4] rely heavily on the use of MFCCs for feeding classifiers like LSTM, CNNs, or k-NN for animal sound classification. In a more technical analysis, standard and enhanced MFCC experiments were further elaborated on by Prabakaran and Sriuppili [7] through certain steps of audio signal analysis that included - pre-emphasis, windowing, FFT, and DCT; compared multiple MFCC-Hybrid configurations. Davis and Mermelstein [8] compared various speech parameterization methods and concluded MFCCs outperform others in recognition accuracy for speech signals. This observation favors the continued dominance of the MFCCs in animal sound classification and warrants their use to proceed with poultry vocalization. Contextual cochleagram features proposed by Sattar [9] beat the MFCCs by over 20% in acoustic recognition performance in the presence of environmental noise on the farms, thus raising concerns about the wide acceptance of MFCCs in smart agriculture settings. Puswal and Liang [10] explored the correlation between vocal features and anatomical traits in chickens. However, while different morphological traits between sexes have been noted, the study has discovered a weak correlation between vocal acoustics and physiology, suggesting behavioral factors and context may have a stronger influence on acoustic variability than morphology. This favors the use of dynamic rather than static acoustic features for classification models in poultry.

Table 1. Comparison of static and dynamic acoustic feature sets in animal vocalization studies. Dynamic features such as cochleagram, SincNet, and wav2vec2 exhibit greater robustness in noisy and real-world farm environments, whereas static features like MFCC and Mel-spectrogram perform well in controlled or low-noise settings.

bermant	Feature Name	Study / Authors	Model Used	Environment	Reported Performance	Notes
Static	MFCC	Umarani et al. [1]	LSTM	General (RAVDESS)	97.22%	Verified via IEEE: LSTM + MFCC for emotion recognition
Static	MFCC	Jung et al. [3]	CNN	General	91.02% (cattle), 75.78% (hens)	Lower for hens—possibly due to background noise
Static	MFCC variants + FFT/DCT	Prabakaran & Sriuppili [7]	MFCC variants	Controlled	94.44%	Comparative setup across MFCC variations

Static	MFCC	Bhandekar et al. [28]	SVM	Lab	95.66%	Strong in low-noise environments
Static	Mel-Spectrogram	Henri et al. [12]	MobileNetV2	Birdsong (natural)	84.21%	Limited context modeling
Dynamic	Cochleagram	Sattar [9]	Context-aware classifier	Noisy farm	>20% higher than MFCC	Better adaptability to environmental noise
Dynamic	SincNet	Bravo Sanchez et al. [52]	Raw waveform classifier	Minimal preprocessing	>65% (NIPS4Bplus)	Learns directly from waveform, robust to distortions
Dynamic	Spectral Entropy	Herborn et al. [18]	Entropy analysis	Chick stress study	Qualitative improvement	Captures emotional states during distress
Dynamic	Wav2vec 2 Embeddings	Swaminathan et al. [26]	Fine-tuned classifier	Real-world bird data	F1 = 0.89	SSL embeddings outperform handcrafted features

The input signals for convolutional networks also often employ spectrograms, especially log-Mel spectrograms. The work of Zhong et al. [11], Henri and Mungloo-Dilmohamud [12], Romero-Mujalli et al. [13], Thomas et al. [14], Mao et al. [5], Mangalam et al. [6], Li et al. [15], and Neethirajan [16] analyzed spectrograms for use in CNNs or spectrogram-based embedding studies. STFT parameters cleanly turned high-quality latent space representations with the help of Mel-scaling and z-normalization, particularly as indicated by Thomas et al. [14] and Sainburg et al. [17].

Spectral entropy is gaining ground as a possible indicator-a-feature for distress. Herborn et al. [18] showed that reduced ratings on the spectral entropy scale of distress calls-from all of which increased calls per day-and long-term welfare and future well-being outcomes in chicks. In the same line, Ginovart-Panisello et al. [19] had fast-induced stress in newly hatched broilers using Butterworth filtered signals and centroid spectral parameters. There are pipelines in a range of past studies to improve preprocessing in real conditions with lots of noise. Tao et al. [20], MFCC, resorted to ZCR and exponential smoothing to filter signals before extracting features. Time masking, SpecSameClassMix and Gaussian noise augmentation were employed to enhance theoretical

robustness of spectrograms in the works of Bermant et al. [21] and Soster et al. [22]. Comprehensive augmentations like frequency masking and noise injection were incorporated as stated by Mao et al. [5]. Thomas et al. [4] included noise suppression layers into their wider strategy for audio cleaning before deep-mould training.

Besides feature transformation, automated segmentation tools have proven efficient, similar to the benchmark ones in Terasaka et al. [23] and Michaud et al. [24]. Such studies involved comparative works using libraries such as Librosa, BirdNET, or Perch and revealed how BirdNET resulted in a higher F1-score. Merino Recalde [25] developed pykanto, the Python library that facilitates semi-automatic segmentation and labeling of large acoustic datasets to use them in deep learning models.

Beyond MFCCs and spectrograms, researchers also seek other acoustic representations. Latent projection techniques were introduced by Sainburg et al. [17], which sidestep traditional hand-crafted features. The importance of embeddings from perusal models trained on raw audio can be illustrated in the work by Swaminathan et al. [26] and Bermant et al. [21]. The representation learned is often superior to the hand-crafted ones. Some studies also use time-domain parameters such as duration, pitch, zero-crossing rate, and energy. For instance, Du et al. [27] extracted nine temporal and spectral features based in source-filter theory to detect thermal discomfort in laying hens. Ginovart-Panisello et al. [34,36–38] often included metrics such as spectral centroid, vocalization rate (VocalNum), and variation in spectral bandwidth in examining the environmental impacts and stress in broiler chickens.

Taken together, these publications show that acoustic feature design is still a very lively arena and a pivotal aspect of poultry vocalization analysis. Feature selection can be completely hand-crafted, learned, or hybrid-the chosen approach affects substantially the robustness and generalizability of the model under the field circumstances of relatively noisy, imbalanced, and unlabeled data.

3. Deep Learning and Classical Models

A vast majority of studies that have analyzed poultry and animal vocalizations concentrate on supervised classification techniques, which range from traditional machine learning models to the latest deep learning architectures. Depending on the aims of the individual projects, data limitations, and computing set up, MFCCs, spectrograms, or combinations of audio representations are trained in the models.

3.1. Classical Machine Learning Models

Some traditional classifiers such as SVM, RF, k-NN, Naive Bayes and Gaussian Naive Bayes have seen their application in the area of poultry sound classification especially in cases of low data and resource-constrained environments. For example, Bhandekar et al. [28] tested four different models (SVM, k-NN, Naive Bayes, and Random Forest) using MFCC features extracted from chicken vocalizations where SVM scored the best with an accuracy of 95.66%. In another example, Pereira et al. [2] reported 85.61% accuracy with a Random Forest model trained on FFT-extracted features to assess the distress of chicks.

Table 2. Performance of classical machine learning models in animal vocalization classification.

Authors	Model(s)	Reported Performance
Pereira et al. [2]	Random Forest	85.61%
Tao et al. [20]	SVM, RF, k-NN	k-NN: 94.16%

Bhandekar et al. [28]	SVM	95.66%
Du et al. [27]	SVM	Sensitivity = 95.1%
Ginovart-Panisello et al. [29]	Gaussian Naive Bayes	F1-score = 80%

Tao et al. [20] considered SVM, RF, CNN, and k-NN for the recognition of broiler vocalizations using multi-domain features, where k-NN eventually achieved the best result with an accuracy of 94.16% after feature selection. Ginovart-Panisello et al. [29] used Gaussian Naive Bayes in detecting vaccine response classified based on MFCCs and spectral centroid, with an F1-score of 80%. Du et al. [27] applied SVMs to temporal-spectral features toward the detection of thermal discomfort at a sensitivity of 95.1%.

3.2. Convolutional Neural Networks (CNNs)

Convolutional Neural Networks (CNNs) have shown tremendous data nowadays in terms of the usage of deep learning architecture for animal vocalization classification. Several studies often apply a standard or customized CNN mechanism onto spectrogram inputs for vocal classification. High performances are realized among birds or poultry via CNNs in vocalization classification by this group of studies including Zhong et al. [11], Henri and Mungloo-Dilmohamud [12], Romero-Mujalli et al. [13], Mao et al. [5], Mangalam et al. [6], and Ginovart-Panisello et al. [29]. Henri and Mungloo-Dilmohamud [12] compared MobileNetV2, InceptionV3, and ResNet50, with MobileNetV2 achieving 84.21% accuracy. According to Mangalam et al. [6], a lightweight custom CNN (i.e., with ~300k parameters) outperformed fine-tuned VGG16. Mao et al. [5] discovered light-VGG11 with a 92.78% decrease in parameters against reference architectures, which retained 95% accuracy. Further, Ginovart-Panisello et al. [29] used CNNs trained on spectrograms for detection of stress. In addition, Mangalam et al. [6], Thomas et al. [4], and Mao et al. [5] demonstrate further contribution regarding the value of CNNs with frozen or fine-tuned pretrained backbones.

Some additional specialized applications are:
- Cuan et al. [77,78]: CNN-based detection of Newcastle disease and avian influenza.
- Ginovart-Panisello et al. [19]: CNNs (ResNet) for detection of acute stress based on vocalization and thermographic data.
- Li et al. [15]: ResNet-50 trained on MFCC+Logfbank features for chick sex detection.

Table 3. Performance of deep learning architectures for animal vocalization classification, including CNN, RNN, and attention-based models.

Authors	Model Type	Reported Performance
Jung et al. [3]	2D CNN	91.02% (cattle), 75.78% (hens)
Mao et al. [5]	Light-VGG11 CNN	95%
Mangalam et al. [6]	Lightweight CNN	92.23%
Romero-Mujalli et al. [13]	DeepSqueak CNN	Detection: 91%, Class: 93%
Henri et al. [12]	MobileNetV2	84.21%
Hu et al. [37]	MFF-ScSEnet CNN	>96%

Gupta et al. [32]	CNN-LMU	Best model
Mousse & Laleye [35]	Attention-based RNN	F1-score = 92.75%
Hassan et al. [34]	Conv1D + Burn Layer	98.55%
Hu et al. [37]	MFF-ScSEnet (attention)	>96%

3.3. Recurrent Models (LSTM, GRU, CRNN)

Research utilizing temporal modeling via RNNs, LSTMs, GRUs, and hybrid CNN-RNN models appear often in literature dealing with the sequential structure of vocalizations. The models were LSTM and GRU-based used for species classification and time-series vocal decoding in Umarani et al. [1] and Bermant et al. [21]. Li et al. [15] and Xu and Chang [31] utilized GRUs and CRNNs to classify health conditions and chick sex. Gupta et al. [32] assessed CNN-LSTM, CNN-GRU, and CNN-LMU over large sets of bird vocalizations with CNN-LMU achieving the best performance. Jung et al. [3] combined CNN with LSTM for vocal classification but reported better performance for 2D ConvNets than for the hybrid model. Huang et al. [33] developed a sequence model to detect poultry feeding behavior based on vocal patterns.

3.4. Hybrid and Attention-Based Architectures

Emerging trends integrating CNNs with attention mechanisms or various architectural innovations have arisen in recent works. A Conv1D-based classifier with Burn Layers (noise-injection modules) was implemented by Hassan et al. [34] to enhance generalization, leading to an impressive accuracy of 98.55%. Mousse and Laleye [35] established an attention-based RNN for hens' behavior recognition and reported an F1 score of 92.75%. Huang et al. [36] proposed ASTNet, a spatio-temporal attention network for video saliency detection which can be adapted for multi-modal poultry monitoring. Hu et al. [37] proposed MFF-ScSEnet, which combines Mel-spectrogram and SincNet features with a squeeze-and-excitation mechanism and more than 96% accuracy over datasets of bird song.

3.5. Performance Benchmarks

Several studies conducted model comparisons: Ginovart-Panisello et al. [29] and Thomas et al. [4] have performed both ablation studies and multi-objective training (classification + age estimation). Bermant et al. [21] benchmarked CNNs and RNNs across echolocation and coda recognition tasks and got over 99% accuracy. Gupta et al. [32] and Ghani et al. [38] conducted studies to judge the model generalization across species and setups, thereby demonstrating the necessity for a training set that is large and varied. Bianco et al. [39] reviewed ML techniques in acoustics, stressing how, when sufficient labeled data is available, data-driven classifiers like SVMs, Neural Networks, and Gaussian Mixtures outperform traditional signal processing-based techniques, and thus weigh the trade-off between model interpretability and classification accuracy-an important consideration in their application for acoustic feature selection and hybrid NLP pipelines along with poultry vocal analysis.

4. Self-Supervised and Transfer Learning Approaches

As there are not many annotated datasets available in the realm of animal vocalization research, transfer learning and self-supervised learning (SSL) have become the methodologies for successfully improving model generalization, reducing training cost, and improving performance when working

under conditions of noise or limited resources. Several studies, mostly focused on poultry and wildlife acoustics, make use of pretrained models, which are commonly developed and fine-tuned for specific species tasks and have been applied on human audio or general bioacoustics.

4.1. Transfer Learning with Pretrained CNNs and Audio Embeddings

Studies have utilized transfer learning through pretraining from large-scale datasets like ImageNet or AudioSet before applying the convolutional model to a novel acoustic signal. Some examples include: Henri and Mungloo-Dilmohamud [12] who refined MobileNetV2, ResNet50 and InceptionV3 for bird song classification, with best accuracy (84.21%) corresponding to MobileNetV2. Thomas et al. [4] transferred PANN (Pretrained Audio Neural Network) weights to a multi-objective CNN for broiler vocalization and age estimation. Mangalam et al. [6] compared a custom CNN with fine-tuned VGG16, concluding that the smaller model worked better under field conditions. Li et al. [15] showed that chick sexing tasks conceived from different architectures (ResNet-50, GRU, CRNN), based on breed and feature type, perform variably. McGinn et al. [40] obtained unsupervised feature embeddings derived from the BirdNET CNN to classify within-species vocalizations, emphasizing its strength without retraining. Ginovart-Panisello et al. [29] applied pretrained CNNs to the spectrograms of hens to induce stress response for vaccinated hens.

4.2. Transformer Models and Speech Pretraining

Vaswani et al. [41] introduced a completely novel architecture in the form of their Transformer- a new architecture that replaces recurrence with multi-head self-attention to parallelize sequence modeling and capture long-range dependencies in the modeling process. It was developed for language tasks, but later became fundamental for many acoustic modeling frameworks, including wav2vec2 and BERT. Its scalability and efficiency even become more crucial for studies on poultry vocalization that require temporal analyses across different contexts. Admittedly, transformers from natural language processing are fast finding utility within audio classification tasks. In a more foundational review concerning AI in livestock, Menezes et al. [42] emphasized the increasing role of transformer-based models and large language models (LLMs) such as BERT and wav2vec2 in agricultural applications. Even though the review mainly covered dairy cattle, it highlights the extent to which such architectures could find application in the study of poultry vocalizations, especially in emotion recognition and welfare prediction. Devlin et al. [43] introduced the new language model; a bidirectional Transformer BERT, trained by means of masked language modeling and next-sentence prediction. Just like many language processing tasks, BERT showed astonishing results in several benchmarks, thereby creating the impetus, in automated response systems, for models such as WHISPER and the fine-tuned version of wav2vec2, which are presently being leveraged for poultry vocalization decoding.

Table 4. Reported performance of transfer learning, self-supervised learning (SSL), and AutoML strategies in animal and bioacoustic vocalization analysis.

Authors	Model / Strategy	Reported Performance
Thomas et al. [4]	PANN + CNN	Balanced Accuracy = 87.9%
Ghani et al. [38]	PaSST (Transformer)	F1 = 0.704
Swaminathan et al. [26]	Fine-tuned wav2vec2	F1 = 0.89
Abzaliev et al. [44]	Pretrained wav2vec2	Outperformed all-frames models

Mørk et al. [51]	Data2Vec SSL	+18% vs. supervised baseline
Bravo Sanchez et al. [52]	SincNet	>65% accuracy
Brydinskyi et al. [53]	Personalized wav2vec2	WER ↓ ~3% (natural), ↓ ~10% (synthetic)
Tosato et al. [54]	AutoKeras NAS (Xception)	Outperformed ResNet, VGG, etc.

Ghani et al. [38] examined transfer learning for large-scale birdsong detection using models like BirdNET and PaSST. The model PaSST, distilled from BirdNET, achieved the highest performance and development in-domain (F1 = 0.704). Swaminathan et al. [26] applied fine-tuning of wav2vec models using bird recordings and a feed-forward classifier against an F1 of 0.89 C-xeno-canto data. Abzaliev et al. [44] used the trained wav2vec2 (on human speech) to classify dog barks in terms of breed, sex, and context categories, outperforming all-frames models. Sarkar and Magimai.-Doss [45] found speech-pretrained SSL models to perform at par with those trained specifically for bioacoustics, making it feasible to reuse human-centric models. Neethirajan [46] studied OpenAI's Whisper model for decoding chicken vocalizations to interpret them semantically in terms of token sequences, which were analyzed then by classifiers of sentiment to deduce the emotional states. Morita et al. [47] used Transformer-based models for long-range dependency studies in Bengalese finch songs: eight syllables appeared to be good context length. Gong et al. [48] introduced the Audio Spectrogram Transformer (AST)-a is convolution-free model that uses patch-based spectrogram inputs fed into a Transformer encoder. AST achieved state-of-the-art accuracy across major audio classification benchmarks, thereby emphasizing the potentiality of attention-based modeling architectures toward structured poultry vocalization analysis.

4.3. Self-Supervised Representation Learning

SSL models have made significant inroads into bioacoustic modeling by reducing the dependency on labeled datasets: Baevski et al. [49] presented wav2vec 2.0, which learns by way of contrastive learning and quantization from raw audio latent representations. It serves as the backbone of several follow-up studies, e.g., [26,44]. Wang et al. [50] applied HuBERT segmenting dog vocalizations and performed grammar induction to discover recurring phone sequences that may reveal meaning in sounds of Canine. Mørk et al. [51] tested Data2Vec-denoising, an approach of robust self-supervised pretraining which can yield up to 18% improvements in accuracy over keyword spotting of supervised baselines. Bravo Sanchez et al. [52] employed SincNet, a neural architecture with parameterized sinc filters for classifying bird vocalizations directly from raw audio waveforms. Attaining more than 65% accuracy on the NIPS4Bplus dataset with minimal preprocessing, this research shows the efficacy of raw-signal-based models for the lower complexity of attack-recognizing classification of poultry vocalizations. In personalized adaptive fine-tuning, Brydinskyi et al. [53] indicated that only 10 minutes of data from an individual could fine-tune wav2vec2 to reduce word error rates: about 3% for natural voices and as much as 10% for synthetic. In personalized adaptive fine-tuning, Brydinskyi et al. [53] indicated that only 10 minutes of data from an individual could fine-tune wav2vec2 to reduce word error rates: about 3% for natural voices and as much as 10% for synthetic.

4.4. AutoML and Neural Architecture Search (NAS)

In addition to the manual transfer learning, some studies employ an active nudging from automated approaches in discovering models: Tosato et al. [54] established an optimal Xception architecture for classifying bird vocalizations by using AutoKeras that is better than MobileNetV2,

ResNet50, and VGG16. Gupta et al. [32] presented the results of exploring a number of deep models on the Cornell Bird Challenge dataset including CNN-LSTM and CNN-LMU, with CNN-LMU achieving the peak accuracy on Red Crossbill calls.

These studies in the aggregate validate the power of pretrained and self-supervised models in enabling accurate, efficient, and scalable animal vocal analysis. Such crossroads include vision-based CNN backbones, language-inspired transformers, or SSL-driven embeddings, where cross-model transfer leads to generalizable, low-data animal sound classification—especially important when annotating precision-livestock contexts since it is often very time-consuming and costly.

5. Emotion, Behavior, and Stress Detection

5.1. Stress Detection via Acoustic Signatures

Well-established evidence exists for stress-related modifications of vocal parameters. One of the very few earlier spectrographic studies on chicken vocalizations was undertaken by Collias and Joos [55], who correlated call types (distress calls, clucking, roosting) with relevant behavioral contexts. They found that calls given with descending frequency were often interpreted as distress calls, whereas those with ascending contours often indicated that they were more pleasurable. This important early study laid the groundwork for behavioral correlate acoustic markers used in avian welfare research. In laying hens, acute stress was detected using a combination of thermographic imaging and CNN-based spectrogram classification by van den Heuvel et al.[30]. This revealed beak and comb temperature reduction and decreased call rate following stressor exposure. In similar fashion, Ginovart-Panisello et al. [19] showed that prolonged fasting caused an alteration of vocalizations in chicks, with call rate (VocalNum) and spectral centroid and bandwidth being significantly altered in comparison to fed controls.

In testing the validity of spectral entropy, Herborn et al. [18] found strong links between entropy and welfare outcomes in the long term (reduced weight gain and increased mortality). Sound calls of domestic chicks during isolation were studied by Collins et al. [56], who related these to various levels of emotional arousal as represented by loudness, frequency, and duration. Lev-Ron et al. [57] taught an artificial neural network to classify responses in vocalizations from broilers subjected to environmental stressors, including cold, heat, and wind. The model accuracy was further enhanced by incorporating variables such as age and waveform length to achieve a mean average precision (mAP) of 0.97. Thus, this approach can be scaled up for stress detection in poultry welfare. The effects of auditory stimuli—including classical music and mechanical noise—were studied by Zhao et al. [58] on fear responses and learning in laying hen chicks. Moderate-level Mozart music exposure caused reduced fearfulness, whereas exposure to high-intensity sound impaired learning and increased stress. The emotional response of hens to their chicks in distress was studied by Edgar et al. [59], who found an increase in heart rate, alertness, and maternal vocalizations of hens when distress was simulated in their chicks by air puffs. This suggests that hens can sense offspring distress and react accordingly, providing support for emotional contagion and further emphasizing the use of vocal cues for welfare inferences in poultry.

5.2. Behavior and Reward-Related Vocalizations

Behavioral responses are mirrored in voice patterns. Zimmerman [60] first worked on the "gakel-call" in hens and established linkages with the emotion of frustration that stems from blocked behaviors. More recently, Zimmerman and Koene [61] demonstrated calls in hens vary depending on the reward anticipated (mealworms, food, substrate), where the frequency shifts of the calls associated with food are related to the expected reward's valence. A human study conducted by McGrath et al. [62] revealed that people could identify the chicken calls reliably associated with rewards, indicating the presence of semantic information encoded within the calls. Neethirajan [63] also studied this topic with the WHISPER model, confirming token-based patterns in chicken distranquil vocalization correlated to emotion. Abzaliev et al. [64], in their turn, analyzed

vocalizations in the Japanese tit (Parus minor), specifically focusing on phoneme structure classification via machine learning that will indeed allow for the differentiation of different call types. The training based on validation with human-labeled data will be the major assist in commissioning and developing a real-time automatic classification system for structured communication in birds. In this regard, such investigations could facilitate the transfer of similar models for detection of poultry call types, for which structured elements may encode important behavioral or emotional states. Schober et al. [65] compiled an extensive and rich acoustic repertoire of Pekin duck vocalizations according to varying stimuli, the sex of the subject, and group configurations. This study applied through statistical methods, including ANOVA, cluster analysis, and canonical discriminant analysis, yielding the identification of 16 distinct vocal types linked to behavioral and environmental contexts. Results demonstrate that vocal diversity and sex-specific patterns can serve as proxies for indicating behavioral correlates, in parallel with call-type variation within poultry.

5.3. Emotion Recognition Models

Emotion decoding has been taking advantage of advanced AI models: Neethirajan [66] reviewed the integration of NLP and sentiment analysis with acoustic sensing for animal emotional detection, proposing hybrid AI systems based on thermographic and vocal inputs. With collaborative annotations by psychologists and veterinarians, Cai et al. [67] developed the DEAL model (Deep Emotional Analysis Learning) to interpret emotional states such as hunger and fear in chickens. Ginovart-Panisello et al. [29] identified post-vaccine anxiety in hens by extracting MFCC and spectral centroid features into a GNB classifier. The classifier obtained an F1-score of 80%, and moreover, experimentally reduced stress during anti-inflammatory treatment. Du et al. [27] reported a strong correlation between thermal distress and squawking/alarm calling in hens (e.g., squawk–THI: R = 0.594), within an SVM setting applied to time-frequency outputs. Gavojdian et al. [68] introduced BovineTalk, a deep-learning explainable ML framework for emotional valence and individuality characterization in dairy cow vocalizations. They reported accuracies of 89.4% for distinguishing high- from low-frequency calls for affective state classification and 72.5% for cow identification using GRU-based models. The methodology has cross-species relevance for poultry emotion recognition either on interpretable acoustic features or spectrogram-based modeling. Lavner and Pérez-Granados [69] underlined emerging techniques in passive acoustic monitoring (PAM) for emotional state estimation, pointing to foundational models and threshold-free density estimation tools.

5.4. Behavioral State and Health Linkages

Not only does behavioral analysis work through sound for emotion, but it also demarcates behavioral activities. With formant structure and pitch-based features, Huang et al. [33] have established a 95% accuracy rate for identifying episodes of eating behavior in chickens. Using attention-based RNNs, Laleye and Mousse [35] classified laying hen behaviors with an F1-score of 92.75%. Fontana et al. [70] found a negative correlation between broiler vocal frequency and weight, thus establishing an association between acoustic cues and physiological growth. Karatsiolis et al. [71] proposed a non-invasive farm monitoring system that uses vocal, visual, and environmental sensor data to-interpret Flock-wide psychological states. Manteuffel et al. [72] reviewed how vocal correlates—like call frequency and formant dispersion—indicate both positive and negative emotional states in multiple species of livestock. Güntürkün [73] reviewed the avian nidopallium caudolaterale (NCL) which, functionally similar to mammalian prefrontal cortex, is involved in decision-making, executive control, and behavioral flexibility. Thus, forming a neuroanatomical basis for understanding poultry vocal behavior complexity, particularly when being stressed—cognitive load, or interest state. Galef and Laland [74] have considered mechanisms of social learning such as imitation and local enhancement across animal species and their contribution to behavioral adaptation and cultural transmission. This provides theoretical justification for researching social influences on vocal behavior in poultry, such as peer-induced stress responses and learned vocal cues. Rugani et al. [75] recorded that 3-day-old chicks possess proto-arithmetic skills, opting for larger

object sets during occlusion-based tests. This early cognitive ability suggests that vocal responses in chicks may encode quantitative or perceptual awareness, further legitimizing studies of poultry behavior that model numeracy-linked vocal characteristics.

5.5. Vocal Indicators of Mental State and Social Emotion

Emotion detection of poultry via vocalization can be meaningfully contextualized using established frameworks such as the Five Domains model (nutrition, environment, health, behavior, and mental state)[112].In particular, vocal measures of distress, anticipation, and contentment correspond to a kind of "Mental State" domain—difficult to quantify objectively, yet accessible for study with machine learning—allowing one to assess emotions without any invasion. These acoustic measures operate to bridge the gap between visible behavior and internal affective states, yielding a more composite view of welfare. From here, we assert that emotional contagion—the affective state of one individual induces a similar affective response in others—has some emergent relevance for poultry welfare studies, with one being that distress calls offered by one chick can raise vocal stress markers in cage mates, indicating a viable emotional space that can be mapped using a group acoustic approach [113]. If such social-affective dynamics could be detected reliably, they may feed into welfare protocols oriented toward interventions at the flock level. Also, convincing evidence emerging from ethology indicates that hens respond differentially to the vocal cues of their chicks, implying maternal empathy. Thus, the possibility exists of quantifying cross-individual emotional synchrony by utilizing acoustic AI to analyze the call-and-response interaction between hens and their chicks. Thus, this opens entirely new avenues for affective computing and animal cognition, stressing the need to now specifically consider how machine learning systems developed for farm animals not only classify individual vocalizations but also discern social and relational emotional cues that seem to become embedded in such vocal interactions.

6. Disease Detection and Health Monitoring

Acoustic analysis is a non-invasive alternative to traditional diagnostics for detecting disease, discomfort, and other mostly physiological anomalies in poultry. Many research studies have employed machine learning models to find health-related vocal markers, to assess disease progression, and to validate the effectiveness of intervention strategies.

6.1. Disease-Specific Detection via Vocal Cues

Specific pathogen vocalization signatures have been identified in various studies including Serbessa et al. who reviewed the clinical syndromes, modes of transmission, and control methods for the most common poultry and pig diseases. This would create an excellent foundation as to interpretation of vocal biomarker correlates for specific health statuses, with comparisons made from different species and disease types. Such baseline would be important in the AI modeling of automated disease detection through vocalization analysis. Cuan et al. [77] proposed a Deep Poultry Vocalisation Network (DPVN) where New Castle disease was identified with 98.5% accuracy through calls of infected to healthy chickens. In a subsequent study Cuan et al.[78] trained a CNN (CSCNN) on spectrograms resulting from avian influenza infected chickens, achieving 97.5% accuracy, with preprocessing including frequency filtering and time-domain augmentation. Xu and Chang also [31] proposed a hybrid model for deep learning fusing vocal and fecal image features for poultry health diagnosis, which gives the highest accuracy compared to single modal models. Neethirajan [46] used Whisper, which took chicken vocalizations and created token sequences that were sentiment-scored to identify emotional states and physiological states. Adebayo et al [79] were able to provide a real-world dataset from over 100 chickens for 65 days. Acoustic changes appeared in untreated birds' calls for 30 days and were often associated with respiratory problems, making it significantly important to establish a baseline for future modeling of disease-related acoustics.

6.2. Physiological Monitoring and Comfort Assessment

Health monitoring also includes assessments of thermal comfort and general wellbeing. Du et al. [27] used spectral features for the prediction of heat stress in hens, which proved to be more than 95% sensitive and could relate the call type to the Temperature-Humidity Index (THI). The study by Li et al. [15] was able to identify chick sex by feature combinations of MFCC, logfbank, and spectrogram across breeds, reporting high accuracy through ResNet-50 and GRU. Puswal and Liang [10] explored the relationship between vocal features and anatomical traits in chickens. The presence of morphological difference based on sex was observable, but it did not display a strong correlation between vocal acoustics and physical traits, indicating behavior and context are likely causes of acoustic variance more than morphology. Thus, it may strengthen dynamic compared to static acoustic features in poultry classification models. He et al. [80] reviewed early detection of diseases by means of sensors and proposed acoustic sensing as one answer that is emerging but underused for monitoring clinical symptoms. Mao et al. [5] made a lightweight convolutional neural network that can monitor in real time the distress of chickens with accuracy above 95% in validation from recordings done in noisy conditions. Soster et al. [22] trained a CNN built from more than 2000 broiler vocalizations in the detection of four call types, including distress calls, achieving a balanced accuracy of 91.1%. Thomas et al. [4] created a dual-objective CNN to classify calls and estimate broiler age, thus showing that the vocal patterns change with development and may indicate health status. ChickenSense, a piezoelectric audio sensing device married to a VGG16 CNN has been developed by Amirivojdan et al. [81] to estimate the feed intake. The model predicted intake at 92% accuracy and a margin of error of ±7%, thus supporting sound proxy for metabolic state.

6.3. Real-World Deployment Considerations

Deployability and robustness form important attributes for practical applications. For instance, implementation of the TinyML model for monitoring chicken health-a highly effective approach under-varied health and environmental conditions-has been demonstrated by Srinivasagan et al. [82] at edge devices. Huang et al. [33] linked vocal changes to physiological states such as hunger and satiety using formant and pitch dynamics to detect feeding behavior.

These studies illustrate the viability of using vocalizations as digital biomarkers for disease, thermal stress, respiratory issues, and overall well-being. Combining bioacoustics with embedded AI models and sensor fusion holds strong promise for continuous, non-invasive health monitoring in poultry farms.

7. Automated Pipelines and Toolkits

The availability of largescale open access bioacoustic data has triggered the need for automated pipelines and toolkits to process, annotate, and analyze vocalizations with little manual effort. In this section, systems and frameworks are discussed that fit into the streamline of data-preprocessing machine-learning pipelines intended for model training and inference in the analysis of animal sounds.

7.1. End-to-End Tools for Bioacoustics

Bioacoustic software tools for automating large parts of the workflow have recently emerged. Gibb et al. [83] described a robust overview of passive acoustic monitoring (PAM) pipelines from sensor hardware to acoustic inference. The role of convolutional neural networks (CNNs), unsupervised clustering, hidden Markov models (HMMs), and cross-correlation techniques have been emphasized for scalable ecological assessment. It also addressed challenges like detection uncertainty, model transferability, and the need for standardized datasets for deployment of automated poultry monitoring systems. Schneider et al. [84] presented the clustering and analysis of sound events (CASE), where 48 clustering methods and audio transformations for animal vocalizations were compared. CASE incorporates windowed, multi-feature extraction and serves as

the benchmarking tool for unsupervised vocal classification. Thomas et al. [14] describes a practical guide that implements Short-Time Fourier Transform (STFT) and Uniform Manifold Approximation and Projection (UMAP) embeddings to build low-dimensional representations of animal calls and gain insights into mislabeling, clustering quality, and interactive visualization. Merino Recalde [25] has developed pykanto, a Python library for large acoustic dataset management. It contains segmentation, semi-supervised labeling, and deep model integration, thus speeding up reproducibility in the pipeline. Nicholson [85] developed Crowsetta, a Python package that converts several annotation formats (e.g. Praat, Audacity, Raven) into a standardized structure, which is compatible with analysis tools like vak and pandas. This interoperability simplifies vocal dataset processing and enhances reproducibility of the analysis across bioacoustic pipelines; hence, it is very beneficial for studies involving different poultry call types. Lapp et al. [86] developed OpenSoundscape, a Python Toolbox for detection, classification, and localization of biological sounds, through a synergy of machine-learning principles and signal processing. BirdSet, presented by Rauch et al. [87], is a large dataset consisting of more than 6800 hours of avian recordings. In that paper, six deep models were benchmarked, and source code is available on Hugging Face to promote reproducibility and model evaluation under covariate shift.
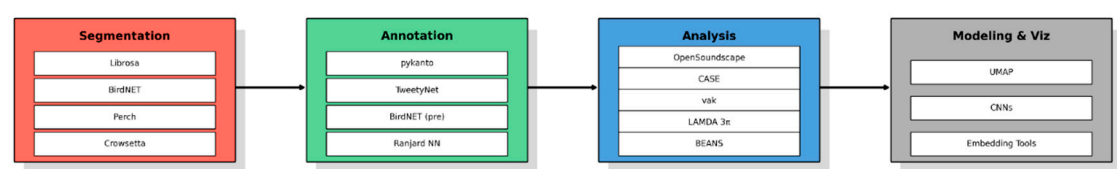

Figure 2. Workflow of Bioacoustic Analysis: Segmentation to Modeling using Specialized Tools.

7.2. Acoustic Segmentation and Dataset Cleaning

For reliable segmentation, high-quality training datasets are essential. In this context, Terasaka et al. [23] compared four segmentation tools in order namely, Librosa, BirdNET, Perch, Few-shot Bioacoustic Event Detection, and concluded that BirdNET was the most accurate. Michaud et al. [24] proposed a DBSCAN and BirdNET-based unsupervised classification method, which ultimately filtered label noise from song datasets thereby enhancing downstream model performance. Sasek et al. [88] introduced a deep supervised source separation (DSSS) framework specialized for site-specific bird vocalization data. A considerable enhancement in separation quality and reduction of downstream labeling errors were achieved by training the ConvTasNet and SuDORMRFNet models using a semi-automated pipeline based on BirdNET, PANNs, and manual filtering. This method shows that integrated pipelines hold great promise when studying poultry calls among other confounding noises in farming settings.

An unsupervised syllable classification approach was developed by Ranjard and Ross [89] with evolving neural networks for the large-scale annotation of bird songs. TweetyNet, a neural network that segments birdsong spectrograms into syllables, was developed by Cohen et al. [90] through end-to-end training, demonstrating good generalizability across species. Lastly, Sethi et al. [91] demonstrated how automated pipelines can scale up biodiversity monitoring by using a BirdNET model pretrained on 152,000+ hours of global audio and manually calibrating detection thresholds for over 100 species.

7.3. Specialized Detection Systems

Lostanlen et al. [92] created BirdVoxDetect (BVD), a freely available system for detecting nocturnal flight calls of birds. It harnesses a multitask CNN to extract features for classification, while

faults in the sensor are detected using random forest model. Michez et al. [93] reported a methodological pipeline using UAS for airborne bioacoustic monitoring of birds and bats. It evaluates drone height and motor noise impacts on call detection rates, with a particular focus on ultra-high frequencies. Their protocol offers a standard for airborne data collection in vocalization-based biodiversity and behavior studies, which may even have further applications in poultry farm surveillance. Guerrero et al. [94] created an unsupervised clustering pipeline (LAMDA 3π) designed for ecological soundscapes. Their approach divides the spectrograms and groups species-specific acoustic clusters (sonotypes), which makes biodiversity assessments possible without labeled data. ChickTrack is the system developed using YOLOv5 plus Kalman filtering in real-time chicken tracking, which is integrated with the monitoring of behaviors using over 3,800 annotated frames from Neethirajan [95]. Bermant et al. [96] present a hybrid pipeline with CNNs for echolocation click detection and RNNs for time-series analysis of sperm whale vocalizations, where transfer learning on proxy tasks allows achieving high-accuracy downstream classification. Berthet et al. [97] reviewed the application of linguistic theory (syntax, semantics, pragmatics) in the animal communication systems and proposed analytical pipelines that include linguistic models into neuroethological data. Hagiwara et al. [98] presented BEANS (Benchmark of Animal Sounds). It is a benchmark that combines 12 different datasets available in public covering birds, mammals, anurans, and insects and sets up classification and detection benchmarks in order to promote standardized evaluation in the field.

These toolkits and pipelines will bring a paradigm shift in the field of animal acoustic analysis, away from individualistic task-specific models toward scalable, generalizable frameworks with standardized data, reproducible pipelines, and automated annotation capacities.

8. On-Farm Deployment and Edge AI

For real-world applications of acoustic monitoring in poultry and livestock, it is essential that machine learning models operate reliably under field conditions. Such system requirements are to be self-sufficient and robust in handling noise and power-efficient operation with low-power edge devices or embedded hardware. All those facts made a strong reflection of the dominant trend in research toward practical and affordable solutions in smart agriculture.

8.1. TinyML and Embedded Inference

With edge AI, mainly through TinyML, real-time inference is performed directly on equipment deployed at farms. In this way, Srinivasagan et al. [82] trained their tiny machine learning models for chicken vocalization using these low-power processors, thus managing memory limitations while maintaining accuracy for multiple health status conditions. The ChickenSense system is a fusion of piezoelectric sensors and the VGG16 model, monitoring the feed intake acoustics of chickens with 92% classification accuracy in +/-7% estimation error (Amirivojdan et al. [81]). Using phase-coding and Gaussian classifiers such as SVM and k-NN on hardware of Raspberry Pi, Bhandekar et al.[28] designed a real-time monitoring system for analysis with synchronized video and audio tracking.Huang et al.[33] developed a module of vocal formants to detect the feeding behavior in noisy field conditions.

Table 5. Comparison of common microphone and acoustic sensor types used in on-farm poultry acoustic monitoring, highlighting trade-offs in signal quality, power, and deployment suitability.

Sensor Type	Example Devices	Sampling Rate	SNR	Power Consumption	Form Factor	Cost (Estimate)	Remarks

Piezoele ctric	ChickenS ense (custom) [81]	16– 44.1 kHz	Mode rate	Very Low	Contact- mount	Low (<$5)	Good for contact- based feeding detectio n
MEMS Microph one	ReSpeak er USB Mic Array	48 kHz	63–72 dB	Low	Beamfor ming array	Moder ate ($25– 40)	Enables directio nal detectio n and active noise cancella tion
Electret Conden ser	Analog mic modules	8–16 kHz	Low– Mid	Moderate	Analog circuit	Very Low (~$2)	Noisy, often used in low- cost setups
MEMS + DSP (digital)	Syntiant NDP101 + mic front-end	16–32 kHz	High	Ultra Low (<1mW)	Edge-ML enabled	Moder ate– High ($40+)	Optimi zed for TinyML & keywor d spotting

TinyML frameworks like TensorFlow Lite Micro, Edge Impulse, and Syntiant now allow optimized models, for example, quantized CNNs or shallow Transformers, to be deployed on low-power microcontrollers such as ARM Cortex-M and ESP32 [115]. Models like these achieve the real-time classification of poultry vocalizations, consuming as little energy as 1-10 mW for continuous monitoring without draining battery-operated IoT systems. In contrast, cloud-based pipelines require constant audio streaming and network bandwidth, which not only increases operational costs but also introduces risks of data leakage, latency bottlenecks, and reliance on external connectivity, particularly problematic in rural farm settings [114]. From an AI systems perspective, edge-AI deployments promise better autonomy and resilience, primarily when combined with local feedback loops that might alert farmers about abnormal distress calls. Yet, how viable edge solutions become is largely dependent on the trust and interpretability underpinning them from the perspective of the farmers. Transparent models with explainable outputs such as call-type labeling and emotion tagging, complemented by local visualization dashboards, will boost the acceptance level of farmers, particularly if privacy-preserving inference methods and fail-safe precautions of the device level are in place.

Table 6. IoT Protocols for Poultry Acoustic + Sensor Monitoring .

Protocol	Range	Bandwidth	Power Efficiency	Cost	Best For	Limitations
LoRaWAN	5–15 km (rural)	Low (0.3–50 kbps)	Excellent	Low to Mod	Long-range farm monitoring	Latency, not for high-frequency data
Zigbee	~10–100 m	Medium (250 kbps)	Good	Low	Local mesh in dense poultry houses	Needs mesh routers, limited range
NB-IoT	1–10 km (urban)	Low–Med (26–127 kbps)	Excellent	Carrier tied	Cellular farms w/ good coverage	Carrier dependency, SIM/data needed
Wi-Fi	~100 m	High (Mbps)	Poor	Moderate	Real-time dashboards & video	Power-hungry, not suitable for edge AI
BLE 5.0	~100–400 m	Low (~2 Mbps)	Excellent	Low	Low-power sensor pairing	Short range, not ideal for big farms

8.2. Robustness to Noise and Uncontrolled Environments

Studies that have addressed the effects of noise and changing environments: Mao et al. [5] employed their lightweight CNN (light-VGG11) for time-continuous recordings and real-farm conditions, confirming its robust performance with over 95% accuracy. Mangalam et al. [6] used on-site smartphone recordings in Indian farms yielding a 92.23% accuracy rate on three vocalizations by using a lightweight CNN. Goyal et al. [99] dealt with systematic review in smart poultry farms, particularly highlighting computer vision, IoT and AI's role in real-time decision support systems and low-cost deployment. Karatsiolis et al. [71] also proposed something similar, where a multi-modal system, vocal and visual environmental sensor models, is designed to perform the assessment of communal flock welfare using a completely non-invasive procedure.

8.3. Sound as a Proxy for Behavior and Environment

Long-term field studies conducted by Ginovart-Panisello et al. [109–111] have illustrated how vocal features (e.g., peak frequency, MFCCs) correlate to temperature, humidity, CO_2 levels, and ventilation conditions across different production cycles. Such studies therefore prove the feasibility of passive acoustic monitoring for environmental assessment and flock health systems. Ginovart-Panisello et al. [29] showed that acoustic responses to vaccination can be automatically tracked under farm conditions, even in the absence of labeled emotional categories. In response to fasting stressors

in commercial hatcheries, Ginovart-Panisello et al. [19] tracked call rates and spectral features in real-time.

Niu et al. [100] reviewed avian visual cognition and associated brain pathways- entopallium and visual Wulst. Their findings corroborate birds' advanced object recognition and tracking capabilities, which provides a neural basis to integrate visual and acoustic signals into behavior monitoring systems. Such integration finds utmost importance in smart poultry surveillance platforms

8.4. Deployment-Friendly Design Practices

Many studies involve optimization to reduce model size, boost energy efficiency, or simplify their architecture:

- Mao et al. [5] reduced the total number of parameters by 92.78% against standard VGG11.
- Hassan et al. [34] introduced Burn Layers (noise-injection modules) to improve generalization under deployment noise.
- Ginovart-Panisello et al. [30] combined thermographic imaging together with CNN-based vocal classifiers to provide an in-field assessment of acute stress in a non-invasive manner.

These studies demonstrate that conjoining edge AI with robust and lightweight architectures is not only possible but a necessity for real deployment in commercial poultry production systems. Continuous monitoring under the decision-making process in a non-invasive and interpretable manner and meeting farm constraints is fast becoming the norm within smart livestock farming.
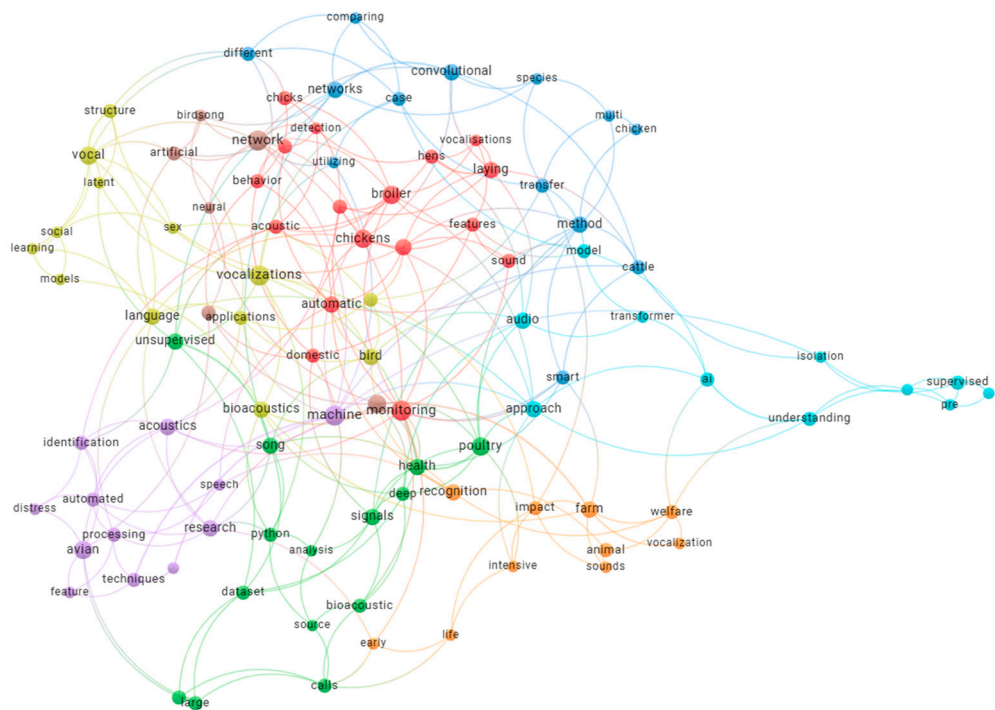

Figure 3. Keyword co-occurrence network showing thematic clusters in livestock vocalization research. Node size indicates keyword frequency, while colors represent distinct research themes such as poultry monitoring, acoustic analysis, and deep learning approaches.

9. Gaps, Challenges, and Future Directions

The current situation in the field of Bioacoustics and analysis leaves an important requirement for further research to identify and fine-tune the limitations associated with reproducibility, generalization, interpretability, and implementation.

9.1. Technical Challenges and Research Gaps

9.1.1. Dataset Limitations and Reproducibility

A heavy emphasis in many studies has been laid on the fact complemented by the presence of few high-quality, and large-sized annotated datasets. Most bioacoustic studies lack full pipeline transparency in their results, as it is usually stated by Mutanu et al. [101]. They recognized qualities of reproducibility in the general consideration of studies, gap in locomotion-related sounds, and inconsistent evaluation metrics being part of systemic issues. As recurring obstacles, Lavner and Pérez-Granados [69] describe low signal-to-noise ratios, class imbalance, and lack of global standardized datasets. Coutant et al. [102] conducted a scoping review of 52 bioacoustic studies across livestock species and identified common acoustic techniques and welfare indicators in this review. Inconsistencies in protocols and an increasing tendency toward ML-driven vocal analysis for automated welfare monitoring were also revealed in this report. This explains the need for standardization in poultry-focused bioacoustics.

9.1.2. Cross-Domain Model Generalization

The question of whether models trained on one species or domain generalize to another is central to future applications. Van Merriënboer et al. [103] reviewed evaluation methods and showed how data variability and covariate shift affect degradation in generalization.Ghani et al. [38] and Gupta et al. [32] showed that transfer learning improves performance, but it still incurs a performance drop in unseen soundscapes or under polyphonic conditions. Swaminathan et al. [26] and Sarkar and Magimai-Doss [45] have shown that self-supervised models pretrained on human speech often outshine those trained from scratch but still require fine-tuning on animal-specific data.

9.1.3. Interpretability and Semantic Representation

Although many have succeeded to high classification rates, the number of works which deal with interpretability of vocal signals is less. Neethirajan [63] and Cai et al. [67] both reached out to semantically decode chicken vocalizations with NLP-inspired models; however, the field has no broadly accepted benchmarks for semantic labeling or emotional annotations. Standard datasets, understandable architectures, and interdisciplinary interactions among acoustics, animal behavior, and machine learning are needed for future research effort, according to Stowell [104].

9.2. Theoretical and Ethical Considerations

9.2.1. Theoretical Foundations and Linguistic Analogues

Bolhuis et al. [105] reject claims of syntactic structure in bird vocalizations, stating that animal communication lacks true combinatorial semantics. Berthet et al. [97] supports the importation of linguistic theories into animal communication (i.e. syntax, pragmatics) and argues that such models should respect certain ethological constraints. Jarvis [106] brought together many lines of research in vocal learning to suggest that animals might share features of language. However, the full accomplishment of vocal learning is rare and biologically constrained.

9.2.2. Theoretical Foundations and Linguistic Analogues

Currently, ethical study is becoming very relevant in AI and animal research. Takeshita and Rzepka [107] identified numerous NLP datasets and models as embedding speciesism, thus warranting the need for the fair representation of nonhuman vocalizations in research and applications. Future studies should be concerned with multimodal systems and their use across a wider range of species. According to Zimmerman [60], Zimmerman and Koene [61], Manteuffel et al. [72], and Marino [108], there is a pressing need for further behavioral and emotional interpretations of poultry vocalizations. Morita et al. [47], Sainburg et al. [17], and Wang et al. [50], extended deep learning for modeling long-range dependencies, latent structures, and grammar-like patterns even in nonhuman species. Cross-species studies like Abzaliev et al. [44], Sethi et al. [91], and Bermant et al.

[96] demonstrated that deep learning pipelines are highly adaptable—but lack interpretability and standardization. The field is moving toward hybrid, explainable, and multi-species-aware models that better bridge computational power with ethological relevance.

9.3. Practical Gaps: Sensor Metrics, IoT Architecture, and Deployment Standards

Despite significant advances in algorithms, the real-world deployment of poultry acoustic AI systems faces practical challenges in sensor evaluation, wireless communication infrastructure, data fusion, and responsible technology design. One major limitation encountered in existing research is the absence of standardized metrics to define microphone and sensor robustness in the presence of a noisy farm environment. Benchmarking in the future should objectively report acoustic performance indicators such as sound-to-noise ratio (SNR), dB(A) ambient noise levels, and the attenuation profile in the frequency bands of interest. Potential techniques for noise cancellation can be explored and applied, for example, through spectral subtraction, Wiener filtering, and neural-based speech enhancement, all promising to improve the performance of the system under heavy noise conditions [116]. The on-farm deployment is also largely dependent on proper selection of the wireless protocols: Technologies have different trade-offs between cost, latency, and energy efficiency. Recently, LoRaWAN has gained attention for its extremely low power consumption and maximum range of 5-15 km, NB-IoT is available as a carrier-integrated medium-bandwidth solution, and Zigbee works in the short range with mesh networking capabilities. For instance, Zigbee is suitable for local mesh needs in densely populated poultry houses, whereas LoRaWAN would provide long-range coverage for widely spaced farms. Those compromises directly affect the system acoustic scale and interoperability and should, therefore, be explicitly considered when planning the infrastructure [117].

Acoustic surveillance systems should comply with both data privacy and sustainability objectives. In the European Union, any system collecting or storing identifiable vocalizations must comply with the General Data Protection Regulation (GDPR) [119]. In parallel, considerations about the rampant deployment of embedded sensors being an electronic waste problem have also emerged. Research now emphasizes sustainable smart farming practices—such as modular sensor designs, recyclable components, and low-power architecture as a means to reduce e-waste and ensure long-term viability [120].

Rare vocalization types—created, for example, to signal the onset of a disease or acute distress—often have limited labeled data. Few-shot learning frameworks, with Prototypical Networks (ProtoNets) being a classical example, provide a way to classify these infrequent events reliably from only very few examples [121]. Inorder to achieve deployment transparency, XAI solutions can be used. For instance, Grad-CAM or LIME visualization techniques [122] can highlight the regions of spectrograms that influence CNN model decision-making, thus helping to boost model trust and, in turn, farmer acceptance. Adoption ultimately hinges on the alignment of the system with a farmer's workflow and usability expectations. Interface formats (e.g., SMS alerts vs. dashboard visualizations), economic modeling (e.g., $50/sensor vs. 10% mortality reduction) and participatory design strategies (e.g., focus groups, usability trials) must be employed for development. Training may be given through applications such as DeepSqueak that will allow farmers and technicians to actively engage in annotation, validation, and deployment, cultivating long-lasting adoption and trust toward the technology.

References

1. Umarani, M., Meyyappan, S., Vallathan, G., & Karthi, G. (2024). LSTM-based vocalization analysis for identification and classification of avian acoustics. Proceedings of the International Conference on Computational Intelligence for Green and Sustainable Technologies (ICCIGST). https://doi.org/10.1109/ICCIGST60741.2024.10717534

2. Pereira, E., Nääs, I.d.A., Ivale, A.H., Garcia, R.G., Lima, N.D.d.S., & Pereira, D.F. (2023). Energy assessment from broiler chicks' vocalization might help improve welfare and production. Animals, 13(1), 15. https://doi.org/10.3390/ani13010015

3. Jung, D.-H., Kim, N. Y., Moon, S. H., Kim, H. S., Lee, T. S., Yang, J.-S., Lee, J. Y., Han, X., & Park, S. H. (2021). Classification of vocalization recordings of laying hens and cattle using convolutional neural network models. Journal of Biosystems Engineering, 46(4), 217–224. https://doi.org/10.1007/s42853-021-00101-1

4. Thomas, P., Grzywalski, T., Hou, Y., de Carvalho, P. S., De Gussem, M., Antonissen, G., & Botteldooren, D. (2023). Using a neural network-based vocalization detector for broiler welfare monitoring. 10th Convention of the European Acoustics Association, Turin, Italy. https://doi.org/10.61782/fa.2023.0474

5. Mao, A., Giraudet, C. S. E., Liu, K., Nolasco, I. D. A., Xie, Z., Xie, Z., ... & McElligott, A. G. (2022). Automated identification of chicken distress vocalizations using deep learning models. Journal of the Royal Society Interface, 19(192), 20210921. https://doi.org/10.1098/rsif.2021.0921

6. Mangalam, K., Sarkar, S., Dogra, Y., Saini, M., & Goel, N. (2024). What Did the Chicken Say: A Multi-class Classification Method on Chicken Vocalizations. In H. N. Saha et al. (Eds.), Proceedings of the International Conference on Systems and Technologies for Smart Agriculture (pp. 667–676). Springer. https://doi.org/10.1007/978-981-97-5157-0_55

7. Prabakaran, D., & Sriuppili, S. (2021). Speech processing: MFCC based feature extraction techniques – An investigation. Journal of Physics: Conference Series, 1717, 012009. https://doi.org/10.1088/1742-6596/1717/1/012009

8. Davis, S., & Mermelstein, P. (1980). Comparison of parametric representations for monosyllabic word recognition in continuously spoken sentences. IEEE Transactions on Acoustics, Speech, and Signal Processing, 28(4), 357–366. https://doi.org/10.1109/TASSP.1980.1163420

9. Sattar, F. (2022). A context-aware method-based cattle vocal classification for livestock monitoring in smart farm. Chem. Proc., 10(89). https://doi.org/10.3390/IO CAG2022-12233

10. Puswal, S. M., & Liang, W. (2024). Acoustic features and morphological parameters of the domestic chickens. Poultry Science, 103(6), 103758. https://doi.org/10.1016/j.psj.2024.103758

11. Zhong, M., Taylor, R., Bates, N., Christey, D., Basnet, H., Flippin, J., ... & Lavista Ferres, J. (2021). Acoustic detection of regionally rare bird species through deep convolutional neural networks. Ecological Informatics, 64, 101333. https://doi.org/10.1016/j.ecoinf.2021.101333

12. Henri, E. J., & Mungloo-Dilmohamud, Z. (2021). A deep transfer learning model for the identification of bird songs: A case study for Mauritius. Proceedings of the International Conference on Electrical, Computer, Communications and Mechatronics Engineering (ICECCME). https://doi.org/10.1109/ICECCME52200.2021.9590917

13. Romero-Mujalli, D., Bergmann, T., Zimmermann, A., & Scheumann, M. (2021). Utilizing DeepSqueak for automatic detection and classification of mammalian vocalizations: a case study on primate vocalizations. Scientific Reports, 11, 24463. https://doi.org/10.1038/s41598-021-03941-1

14. Thomas, M., Jensen, F. H., Averly, B., Demartsev, V., Manser, M. B., Sainburg, T., & Roch, M. A. (2022). A practical guide for generating unsupervised, spectrogram-based latent space representations of animal vocalizations. Journal of Animal Ecology, 91(8), 1567–1581. https://doi.org/10.1111/1365-2656.13754

15. Li, Z., Zhang, T., Cuan, K., Fang, C., Zhao, H., Guan, C., Yang, Q., & Qu, H. (2022). Sex detection of chicks based on audio technology and deep learning methods. Animals, 12(23), 3106. https://doi.org/10.3390/ani12223106

16. Neethirajan, S. (2023). Vocalization patterns in laying hens—An analysis of stress-induced audio responses. bioRxiv. https://doi.org/10.1101/2023.12.26.573338

17. Sainburg, T., Thielk, M., & Gentner, T. Q. (2020). Finding, visualizing, and quantifying latent structure across diverse animal vocal repertoires. PLoS Computational Biology, 16(10), e1008228. https://doi.org/10.1371/journal.pcbi.1008228

18. Herborn, K. A., McElligott, A. G., Mitchell, M. A., Sandilands, V., Bradshaw, B., & Asher, L. (2020). Spectral entropy of early-life distress calls as an iceberg indicator of chicken welfare. Journal of the Royal Society Interface, 17(167), 20200086. https://doi.org/10.1098/rsif.2020.0086

19. Ginovart-Panisello, G. J., Iriondo, I., Panisello Monjo, T., Riva, S., Garcia, R., Valls, J., & Alsina-Pagès, R. M. (2024). Acoustic detection of the effects of prolonged fasting on newly hatched broiler chickens. Computers and Electronics in Agriculture, 219, 108763. https://doi.org/10.1016/j.compag.2024.108763

20. Tao, W., Wang, G., Sun, Z., Xiao, S., Wu, Q., & Zhang, M. (2022). Recognition method for broiler sound signals based on multi-domain sound features and classification model. Sensors, 22(20), 7935. https://doi.org/10.3390/s22207935

21. Bermant, P. C., Bronstein, M. M., Wood, R. J., Gero, S., & Gruber, D. F. (2019). Deep machine learning techniques for the detection and classification of sperm whale bioacoustics. Scientific Reports, 9, 12588. https://doi.org/10.1038/s41598-019-48909-4

22. Soster, P., Grzywalski, T., Hou, Y., Thomas, P., Dedeurwaerder, A., De Gussen, M., Tuyttens, F., Devos, P., Botteldooren, D., & Antonissen, G. (2024). A machine learning approach for broiler chicken vocalization monitoring. SSRN Electronic Journal. https://doi.org/10.2139/ssrn.4999753

23. Terasaka, D. T., Martins, L. E., dos Santos, V. A., Ventura, T. M., de Oliveira, A. G., & Pedroso, G. S. G. (2024). Bird audio segmentation: Audio segmentation to build bird training datasets. Anais Estendidos do Workshop de Computação Aplicada à Gestão do Meio Ambiente e Recursos Naturais (WCAMA), 22–29. https://doi.org/10.5753/wcama.2024.2055

24. Michaud, F., Sueur, J., Le Cesne, M., & Haupert, S. (2023). Unsupervised classification to improve the quality of a bird song recording dataset. Ecological Informatics, 74, 101952. https://doi.org/10.1016/j.ecoinf.2022.101952

25. Merino Recalde, N. (2023). pykanto: A python library to accelerate research on wild bird song. Methods in Ecology and Evolution, 14(8), 1994–2002. https://doi.org/10.1111/2041-210X.14155

26. Swaminathan, B., Jagadeesh, M., & Subramaniyaswamy, V. (2024). Multi-label classification for acoustic bird species detection using transfer learning approach. Ecological Informatics, 80, 102471. https://doi.org/10.1016/j.ecoinf.2024.102471

27. Du, X., Carpentier, L., Teng, G., Liu, M., Wang, C., & Norton, T. (2020). Assessment of laying hens' thermal comfort using sound technology. Sensors, 20(2), 473. https://doi.org/10.3390/s20020473

28. Bhandekar, A., Udutalapally, V., & Das, D. (2023). Acoustic based chicken health monitoring in smart poultry farms. 2023 IEEE International Symposium on Smart Electronic Systems (iSES), 224–229. https://doi.org/10.1109/iSES58672.2023.00054

29. Ginovart-Panisello, G. J., Iriondo, I., Panisello Monjo, T., Riva, S., Cancer, J. C., & Alsina-Pagès, R. M. (2024). Acoustic detection of vaccine reactions in hens for assessing anti-inflammatory product efficacy. Applied Sciences, 14, 2156. https://doi.org/10.3390/app14052156

30. van den Heuvel, H., Youssef, A., Grat, L. M., & Neethirajan, S. (2022). Quantifying the effect of an acute stressor in laying hens using thermographic imaging and vocalisations. bioRxiv. https://doi.org/10.1101/2022.07.31.502171

31. Xu, R.-Y., & Chang, C.-L. (2024). Deep Learning-Based Poultry Health Diagnosis: Detecting Abnormal Feces and Analyzing Vocalizations. 2024 10th International Conference on Applied System Innovation (ICASI), 55–57. https://doi.org/10.1109/ICASI60819.2024.10547723

32. Gupta, G., Kshirsagar, M., Zhong, M., Gholami, S., & Lavista Ferres, J. (2021). Comparing recurrent convolutional neural networks for large scale bird species classification. Scientific Reports, 11, 17085. https://doi.org/10.1038/s41598-021-96446-w

33. Huang, J., Zhang, T., Cuan, K., & Fang, C. (2021). An intelligent method for detecting poultry eating behaviour based on vocalization signals. Computers and Electronics in Agriculture, 180, 105884. https://doi.org/10.1016/j.compag.2020.105884

34. Hassan, E., Elbedwehy, S., Shams, M. Y., ElHafeez, T. A., & ElRashidy, N. (2024). Optimizing poultry audio signal classification with deep learning and burn layer fusion. Journal of Big Data, 11, 135. https://doi.org/10.1186/s40537-024-00985-8

35. Laleye, F. A. A., & Mousse, M. A. (2024). Attention-based recurrent neural network for automatic behavior laying hen recognition. Multimedia Tools and Applications, 83, 62443–62458. https://doi.org/10.1007/s11042-024-18241-9

36. Huang, L., Yan, P., Li, G., Wang, Q., & Lin, L. (2019). Attention Embedded Spatio-Temporal Network for Video Salient Object Detection. IEEE Access, 7, 166203–166213. https://doi.org/10.1109/ACCESS.2019.2953046

37. Hu, S., Chu, Y., Wen, Z., Zhou, G., Sun, Y., & Chen, A. (2023). Deep learning bird song recognition based on MFF-ScSEnet. Ecological Indicators, 154, 110844. https://doi.org/10.1016/j.ecolind.2023.110844

38. Ghani, B., Kalkman, V. J., Planqué, B., Vellinga, W.-P., Gill, L., & Stowell, D. (2024). Generalization in birdsong classification: Impact of transfer learning methods and dataset characteristics. Frontiers in Bird Science, 3, 1515383. https://doi.org/10.48550/arXiv.2409.15383

39. Bianco, M. J., Gerstoft, P., Traer, J., Ozanich, E., Roch, M. A., Gannot, S., & Deledalle, C.-A. (2019). Machine learning in acoustics: Theory and applications. The Journal of the Acoustical Society of America, 146(5), 3590–3628. https://doi.org/10.1121/1.5133944

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

41. Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, Ł., & Polosukhin, I. (2017). Attention is all you need. arXiv preprint arXiv:1706.03762. https://doi.org/10.48550/arXiv.1706.03762

42. Menezes, G. L., Mazon, G., Ferreira, R. E. P., Cabrera, V. E., & Dorea, J. R. R. (2024). Artificial intelligence for livestock: a narrative review of the applications of computer vision systems and large language models for animal farming. Animal Frontiers, 14(6). https://doi.org/10.1093/af/vfae048

43. Devlin, J., Chang, M.-W., Lee, K., & Toutanova, K. (2019). BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers), 4171–4186. https://doi.org/10.18653/v1/N19-1423

44. Abzaliev, A., Pérez Espinosa, H., & Mihalcea, R. (2024). Towards Dog Bark Decoding: Leveraging Human Speech Processing for Automated Bark Classification. arXiv preprint arXiv:2404.18739. https://doi.org/10.48550/arXiv.2404.18739

45. Sarkar, E., & Magimai.-Doss, M. (2025). Comparing Self-Supervised Learning Models Pre-Trained on Human Speech and Animal Vocalizations for Bioacoustics Processing. arXiv preprint arXiv:2501.05987. https://doi.org/10.48550/arXiv.2501.05987

46. Neethirajan, S. (2025). Adapting a Large-Scale Transformer Model to Decode Chicken Vocalizations: A Non-Invasive AI Approach to Poultry Welfare. AI, 6(4), 65. https://doi.org/10.3390/ai6040065

47. Morita, T., Koda, H., Okanoya, K., & Tachibana, R. O. (2021). Measuring context dependency in birdsong using artificial neural networks. PLoS Computational Biology, 17(12), e1009707. https://doi.org/10.1371/journal.pcbi.1009707

48. Gong, Y., Chung, Y.-A., & Glass, J. (2021). AST: Audio Spectrogram Transformer. arXiv preprint arXiv:2104.01778. https://doi.org/10.48550/arXiv.2104.01778

49. Baevski, A., Zhou, H., Mohamed, A., & Auli, M. (2020). wav2vec 2.0: A Framework for Self-Supervised Learning of Speech Representations. arXiv preprint arXiv:2006.11477. https://doi.org/10.48550/arXiv.2006.11477

50. Wang, T. S., Li, X., Zhang, C., Wu, M., & Zhu, K. Q. (2024). Phonetic and Lexical Discovery of Canine Vocalization. In Findings of the Association for Computational Linguistics: EMNLP 2024 (pp. 13972–13983). Association for Computational Linguistics. https://doi.org/10.18653/v1/2024.findings-emnlp.816

51. Mørk, J., Bovbjerg, H. S., Kiss, G., & Tan, Z.-H. (2024). Noise-robust keyword spotting through self-supervised pretraining. arXiv preprint arXiv:2403.18560. https://doi.org/10.48550/arXiv.2403.18560

52. Bravo Sanchez, F. J., Hossain, R., English, N. B., & Moore, S. T. (2021). Bioacoustic classification of avian calls from raw sound waveforms with an open-source deep learning architecture. Scientific Reports, 11, 15733. https://doi.org/10.1038/s41598-021-95076-6

53. Brydinskyi, V., Sabodashko, D., Khoma, Y., Podpora, M., Konovalov, A., & Khoma, V. (2024). Enhancing automatic speech recognition with personalized models: Improving accuracy through individualized fine-tuning. IEEE Access, 12, 116649–116656. https://doi.org/10.1109/ACCESS.2024.3443811

54. Tosato, G., Shehata, A., Janssen, J., Kamp, K., Jati, P., & Stowell, D. (2023). Auto deep learning for bioacoustic signals. arXiv preprint arXiv:2311.04945v2. https://arxiv.org/abs/2311.04945

55. Collias, N., & Joos, M. (1953). The spectrographic analysis of sound signals of the domestic fowl. Behaviour, 5(3), 175–188. https://www.jstor.org/stable/4532776

56. Collins, S. A., Herborn, K., Sufka, K. J., Asher, L., & Brilot, B. (2024). Do I sound anxious? Emotional arousal is linked to changes in vocalisations in domestic chicks (Gallus gallus dom.). Applied Animal Behaviour Science, 277, 106359. https://doi.org/10.1016/j.applanim.2024.106359

57. Lev-Ron, T., Yitzhaky, Y., Halachmi, I., & Druyan, S. (2025). Classifying vocal responses of broilers to environmental stressors via artificial neural network. Animal, 19(1), 101378. https://doi.org/10.1016/j.animal.2024.101378

58. Zhao, S., Cui, W., Yin, G., Wei, H., Li, J., & Bao, J. (2023). Effects of different auditory environments on behavior, learning ability, and fearfulness in 4-week-old laying hen chicks. Animals, 13(19), 3022. https://doi.org/10.3390/ani13193022

59. Edgar, J. L., Lowe, J. C., Paul, E. S., & Nicol, C. J. (2011). Avian maternal response to chick distress. Proceedings of the Royal Society B: Biological Sciences, 278(1721), 3129–3134. https://doi.org/10.1098/rspb.2010.2701

60. Zimmerman, P. H. (1999). "Say what?" Vocalisation as an indicator of welfare in the domestic laying hen (Doctoral dissertation, Wageningen University). ISBN 90-5808-159-1.

61. McGrath, N., Dunlop, R., Dwyer, C., Burman, O., & Phillips, C. J. C. (2017). Hens vary their vocal repertoire and structure when anticipating different types of reward. Animal Behaviour, 128, 79–86. https://doi.org/10.1016/j.anbehav.2017.05.025

62. McGrath, N., Phillips, C. J. C., Burman, O. H. P., Dwyer, C. M., & Henning, J. (2024). Humans can identify reward-related call types of chickens. Royal Society Open Science, 11(1), 231284. https://doi.org/10.1098/rsos.231284

63. Neethirajan, S. (2024). Decoding the Language of Chickens - An Innovative NLP Approach to Enhance Poultry Welfare. bioRxiv. https://doi.org/10.1101/2024.04.29.591707

64. Abzaliev, A., Ibaraki, K., Shibata, K., & Mihalcea, R. (2024). Vocalizations of the Parus minor Bird: Taxonomy and Automatic Classification. In Proceedings of the International Conference on Animal-Computer Interaction (ACI 2024). Association for Computing Machinery. https://doi.org/10.1145/3702336.3702344

65. Schober, J. M., Merritt, J., Ulrey, M., Yap, T. Y., Lucas, J. R., & Fraley, G. S. (2024). Vocalizations of the Pekin duck (Anas platyrhynchos domesticus): How stimuli, sex, and social groups affect their vocal repertoire. Poultry Science, 103, 103738. https://doi.org/10.1016/j.psj.2024.103738

66. Neethirajan, S. (2024). From predictive analytics to emotional recognition – The evolving landscape of cognitive computing in animal welfare. International Journal of Cognitive Computing in Engineering, 5, 123–131. https://doi.org/10.1016/j.ijcce.2024.02.003

67. Cai, J., Yan, Y., & Cheok, A. (2023). Deciphering Avian Emotions: A Novel AI and Machine Learning Approach to Understanding Chicken Vocalizations. Research Square. https://doi.org/10.21203/rs.3.rs-3034567/v1

68. Gavojdian, D., Mincu, M., Lazebnik, T., Oren, A., Nicolae, I., & Zamansky, A. (2024). BovineTalk: Machine learning for vocalization analysis of dairy cattle under the negative affective state of isolation. Frontiers in Veterinary Science, 11, 1357109. https://doi.org/10.3389/fvets.2024.1357109

69. Lavner, Y., & Pérez-Granados, C. (2024). Editorial: Computational bioacoustics and automated recognition of bird vocalizations: New tools, applications and methods for bird monitoring. Frontiers in Bird Science, 3, 1518077. https://doi.org/10.3389/fbirs.2024.1518077

70. Fontana, I., Tullo, E., Butterworth, A., & Guarino, M. (2015). An innovative approach to predict the growth in intensive poultry farming. Computers and Electronics in Agriculture, 119, 178–183. https://doi.org/10.1016/j.compag.2015.10.001

71. Karatsiolis, S., Panagi, P., Vassiliades, V., Kamilaris, A., Nicolaou, N., & Stavrakis, E. (2024). Towards understanding animal welfare by observing collective flock behaviors via AI-powered analytics. Annals of Computer Science and Information Systems, 39, 643–648. https://doi.org/10.15439/2024F2064

72. Manteuffel, G., Puppe, B., & Schön, P. C. (2004). Vocalization of farm animals as a measure of welfare. Applied Animal Behaviour Science, 88(1–2), 163–182. https://doi.org/10.1016/j.applanim.2004.02.012

73. Güntürkün, O. (2005). The avian 'prefrontal cortex' and cognition. Current Opinion in Neurobiology, 15(6), 686–693. https://doi.org/10.1016/j.conb.2005.10.003

74. Galef, B. G., & Laland, K. N. (2005). Social learning in animals: Empirical studies and theoretical models. BioScience, 55(6), 489–499. https://doi.org/10.1641/0006-3568(2005)055[0489:SLIAES]2.0.CO;2

75. Rugani, R., Fontanari, L., Simoni, E., Regolin, L., & Vallortigara, G. (2009). Arithmetic in newborn chicks. Proceedings of the Royal Society B: Biological Sciences, 276(1666), 2451–2460. https://doi.org/10.1098/rspb.2009.0044

76. Serbessa, T. A., Geleta, Y. G., & Terfa, I. O. (2023). Review on diseases and health management of poultry and swine. International Journal of Avian & Wildlife Biology, 7(1), 27–38. https://doi.org/10.15406/ijawb.2023.07.00187

77. Cuan, K., Zhang, T., Li, Z., Huang, J., Ding, Y., & Fang, C. (2022). Automatic Newcastle disease detection using sound technology and deep learning method. Computers and Electronics in Agriculture, 194, 106740. https://doi.org/10.1016/j.compag.2022.106740

78. Cuan, K., Zhang, T., Huang, J., Fang, C., & Guan, Y. (2020). Detection of avian influenza-infected chickens based on a chicken sound convolutional neural network. Computers and Electronics in Agriculture, 178, 105688. https://doi.org/10.1016/j.compag.2020.105688

79. Adebayo, S., Aworinde, H. O., Akinwunmi, A. O., Alabi, O. M., Ayandiji, A., Sakpere, A. B., Adeyemo, A., Oyebamiji, A. K., Olaide, O., & Kizito, E. (2023). Enhancing poultry health management through machine learning-based analysis of vocalization signals dataset. Data in Brief, 50, 109528. https://doi.org/10.1016/j.dib.2023.109528

80. He, P., Chen, Z., Yu, H., Hayat, K., He, Y., Pan, J., & Lin, H. (2022). Research Progress in the Early Warning of Chicken Diseases by Monitoring Clinical Symptoms. Applied Sciences, 12(11), 5601. https://doi.org/10.3390/app12115601

81. Amirivojdan, A., Nasiri, A., Zhou, S., Zhao, Y., & Gan, H. (2024). ChickenSense: A Low-Cost Deep Learning-Based Solution for Poultry Feed Consumption Monitoring Using Sound Technology. AgriEngineering, 6(3), 2115–2129. https://doi.org/10.3390/agriengineering6030124

82. Srinivasagan, R., El Sayed, M. S., Al-Rasheed, M. I., & Alzahrani, A. S. (2025). Edge intelligence for poultry welfare: Utilizing tiny machine learning neural network processors for vocalization analysis. PLOS ONE, 20(1), e0316920. https://doi.org/10.1371/journal.pone.0316920

83. Gibb, R., Browning, E., Glover-Kapfer, P., & Jones, K. E. (2019). Emerging opportunities and challenges for passive acoustics in ecological assessment and monitoring. Methods in Ecology and Evolution, 10(2), 169–185. https://doi.org/10.1111/2041-210X.13101

84. Schneider, S., Hammerschmidt, K., & Dierkes, P.W. (2022). Introducing the Software CASE (Cluster and Analyze Sound Events) by Comparing Different Clustering Methods and Audio Transformation Techniques Using Animal Vocalizations. Animals, 12(16), 2020. https://doi.org/10.3390/ani12162020

85. Nicholson, D. (2023). Crowsetta: A Python tool to work with any format for annotating animal vocalizations and bioacoustics data. Journal of Open Source Software, 8(84), 5338. https://doi.org/10.21105/joss.05338

86. Lapp, S., Rhinehart, T., Freeland-Haynes, L., Khilnani, J., Syunkova, A., & Kitzes, J. (2023). OpenSoundscape: An open-source bioacoustics analysis package for Python. Methods in Ecology and Evolution, 14(9), 2321–2328. https://doi.org/10.1111/2041-210X.14196

87. Rauch, L., Schwinger, R., Wirth, M., Heinrich, R., Huseljic, D., Herde, M., Lange, J., Kahl, S., Sick, B., Tomforde, S., & Scholz, C. (2025). BirdSet: A Large-Scale Dataset for Audio Classification in Avian Bioacoustics. arXiv preprint arXiv:2403.10380. https://doi.org/10.48550/arXiv.2403.10380

88. Sasek, J., Allison, B., Contina, A., Knobles, D., Wilson, P., & Keitt, T. (2024). Semiautomated generation of species-specific training data from large, unlabeled acoustic datasets for deep supervised birdsong isolation. PeerJ, 12, e17854. https://doi.org/10.7717/peerj.17854

89. Ranjard, L., & Ross, H. A. (2008). Unsupervised bird song syllable classification using evolving neural networks. The Journal of the Acoustical Society of America, 123(6), 4358–4368. https://doi.org/10.1121/1.2903861

90. Cohen, Y., Nicholson, D. A., Sanchioni, A., Mallaber, E. K., Skidanova, V., & Gardner, T. J. (2022). Automated annotation of birdsong with a neural network that segments spectrograms. eLife, 11, e63853.. https://doi.org/10.7554/eLife.63853

91. Sethi, S. S., Bick, A., Chen, M.-Y., Crouzeilles, R., Hillier, B. V., Lawson, J., . & Banks-Leite, C. (2024). Large-scale avian vocalization detection delivers reliable global biodiversity insights. Proceedings of the National Academy of Sciences, 121(33), e2315933121. https://doi.org/10.1073/pnas.2315933121

92. Lostanlen, V., Cramer, A., Salamon, J., Farnsworth, A., Van Doren, B. M., Kelling, S., & Bello, J. P. (2024). BirdVoxDetect: Large-Scale Detection and Classification of Flight Calls for Bird Migration Monitoring. IEEE/ACM Transactions on Audio, Speech, and Language Processing, 32, 4134–4145. https://doi.org/10.1109/TASLP.2024.3444486

93. Michez, A., Broset, S., & Lejeune, P. (2021). Ears in the sky: Potential of drones for the bioacoustic monitoring of birds and bats. Drones, 5(1), 9. https://doi.org/10.3390/drones5010009

94. Guerrero, M. J., Bedoya, C. L., López, J. D., Daza, J. M., & Isaza, C. (2023). Acoustic animal identification using unsupervised learning. Methods in Ecology and Evolution, 14(6), 1500–1514. https://doi.org/10.1111/2041-210X.14103

95. Neethirajan, S. (2022). ChickTrack – A quantitative tracking tool for measuring chicken activity. Measurement, 191, 110819. https://doi.org/10.1016/j.measurement.2022.110819

96. Bermant, P. C., Bronstein, M. M., Wood, R. J., Gero, S., & Gruber, D. F. (2019). Deep machine learning techniques for the detection and classification of sperm whale bioacoustics. Scientific Reports, 9, 12588. https://doi.org/10.1038/s41598-019-48909-4

97. Berthet, M., Coye, C., Dezecache, G., & Kuhn, J. (2023). Animal linguistics: a primer. Biological Reviews, 98(1), 81–98. https://doi.org/10.1111/brv.12897

98. Hagiwara, M., Hoffman, B., Liu, J.-Y., Cusimano, M., Effenberger, F., & Zacarian, K. (2022). BEANS: The Benchmark of Animal Sounds. arXiv preprint arXiv:2210.12300. https://doi.org/10.48550/arXiv.2210.12300

99. Goyal, V., Yadav, A., & Mukherjee, R. (2024). A Literature Review on the Role of Internet of Things, Computer Vision, and Sound Analysis in a Smart Poultry Farm. ACS Agricultural Science & Technology, 4(4), 368–388. https://doi.org/10.1021/acsagscitech.3c00467

100. Niu, X., Jiang, Z., Peng, Y., Huang, S., Wang, Z., & Shi, L. (2022). Visual cognition of avians and its underlying neural mechanism: a review. Avian Research, 13, 100023. https://doi.org/10.1016/j.avrs.2022.100023

101. Mutanu, L., Gohil, J., Gupta, K., Wagio, P., & Kotonya, G. (2022). A review of automated bioacoustics and general acoustics classification research. Sensors, 22(21), 8361. https://doi.org/10.3390/s22218361

102. Coutant, M., Villain, A. S., & Briefer, E. F. (2024). A scoping review of the use of bioacoustics to assess various components of farm animal welfare. Applied Animal Behaviour Science, 275, 106286. https://doi.org/10.1016/j.applanim.2024.106286

103. Van Merriënboer, B., Hamer, J., Dumoulin, V., Triantafillou, E., & Denton, T. (2024). Birds, bats and beyond: evaluating generalization in bioacoustics models. Frontiers in Bird Science, 3, 1369756. https://doi.org/10.3389/fbirs.2024.1369756

104. Stowell, D. (2022). Computational bioacoustics with deep learning: a review and roadmap. PeerJ, 10, e13152. https://doi.org/10.7717/peerj.13152

105. Bolhuis, J. J., Beckers, G. J. L., Huybregts, M. A. C., Berwick, R. C., & Everaert, M. B. H. (2018). Meaningful syntactic structure in songbird vocalizations? PLOS Biology, 16(6), e2005157. https://doi.org/10.1371/journal.pbio.2005157

106. Jarvis, E. D. (2019). Evolution of vocal learning and spoken language. Science, 366(6461), 50–54. https://doi.org/10.1126/science.aax0287

107. Takeshita, M., & Rzepka, R. (2024). Speciesism in natural language processing research. AI and Ethics. https://doi.org/10.1007/s43681-024-00606-3

108. Marino, L. (2017). Thinking chickens: a review of cognition, emotion, and behavior in the domestic chicken. Animal Cognition, 20(2), 127–147. https://doi.org/10.1007/s10071-016-1064-4

109. Ginovart-Panisello, G. J., Alsina-Pagès, R. M., Iriondo Sanz, I., Panisello Monjo, T., & Call Prat, M. (2020). Acoustic description of the soundscape of a real-life intensive farm and its impact on animal welfare: A preliminary analysis of farm sounds and bird vocalisations. Sensors, 20(17), 4732. https://doi.org/10.3390/s20174732

110. Ginovart-Panisello, G.-J., Alsina-Pagès, R. M., & Panisello Monjo, T. (2020). Acoustic description of bird broiler vocalisations in a real-life intensive farm and its impact on animal welfare: A comparative analysis of recordings. Engineering Proceedings, 2(1), 53. https://doi.org/10.3390/ecsa-7-08164

111. Ginovart-Panisello, G. J., Iriondo Sanz, I., Panisello Monjo, T., Riva, S., Garriga Dicuzzo, T., Abancens Escuer, E., & Alsina-Pagès, R. M. (2022). Trend and representativeness of acoustic features of broiler chicken vocalisations related to CO_2. Applied Sciences, 12(20), 10480. https://doi.org/10.3390/app122010480

112. Mellor DJ, Beausoleil NJ, Littlewood KE, McLean AN, McGreevy PD, Jones B, Wilkins C. The 2020 Five Domains Model: Including Human-Animal Interactions in Assessments of Animal Welfare. Animals (Basel). 2020 Oct 14;10(10):1870. https://doi.org/10.3390/ani10101870

113. Herrando, C., & Constantinides, E. (2021). Emotional contagion: A brief overview and future directions. Frontiers in Psychology, 12, Article 712606. https://doi.org/10.3389/fpsyg.2021.712606

114. Sanchez-Iborra R, Zoubir A, Hamdouchi A, Idri A, Skarmeta A. Intelligent and Efficient IoT Through the Cooperation of TinyML and Edge Computing. Informatica. 2023;34(1):147-168. https://doi.org/10.15388/22-INFOR505

115. Banbury, C. R., Reddi, V. J., Lam, M., Fu, W., Fazel, A., Holleman, J., Huang, X., Hurtado, R., Kanter, D., Lokhmotov, A., Patterson, D., Pau, D., Seo, J., Sieracki, J., Thakker, U., Verhelst, M., & Yadav, P. (2020). Benchmarking TinyML Systems: Challenges and Direction. arXiv. https://doi.org/10.48550/arXiv.2003.04821

116. Loizou, P.C. (2013). Speech Enhancement: Theory and Practice, Second Edition (2nd ed.). CRC Press. https://doi.org/10.1201/b14529

117. Sinha, R. S., Wei, Y., & Hwang, S. H. (2017). A survey on LPWA technology: LoRa and NB-IoT. ICT Express, 3(1), 14–21. https://doi.org/10.1016/j.icte.2017.03.004

118. European Parliament and Council. (2016). Regulation (EU) 2016/679 (General Data Protection Regulation). Official Journal of the European Union, L119. https://eur-lex.europa.eu/eli/reg/2016/679/oj

119. Gemtou, M., Casares Guillén, B., & Anastasiou, E. (2024). Smart farming technologies and sustainability. In T. Lynn, J. G. Mooney, & P. Rosati (Eds.), Digital Sustainability (pp. 99–120). Palgrave Macmillan. https://doi.org/10.1007/978-3-031-61749-2_6

120. Snell, J., Swersky, K., & Zemel, R. S. (2017). Prototypical networks for few-shot learning. Advances in Neural Information Processing Systems, 30. https://github.com/jakesnell/prototypical-networks

121. Selvaraju, R. R. et al. (2017). Grad-CAM: Visual explanations from deep networks via gradient-based localization. ICCV, 618–626. https://doi.org/10.1109/ICCV.2017.74