DECODING POULTRY WELFARE: AN INTEGRATED MACHINE LEARNING AND NLP FRAMEWORK FOR VOCALIZATION ANALYSIS

By

Venkatraman Manikandan

Submitted in partial fulfilment of the requirements

for the degree of Master of Computer Science

at

Dalhousie University,

Halifax, Nova Scotia

October 2025

© Copyright Venkatraman Manikandan, 2025

For my ***Grandfather****,*

Chathapuram Vaithyanathan Venkataraman

and my ***Extended Family***

Thank you for all the support and love you had showered upon me.

Table of Contents

List of Tables

List of Figures

Abstract

Poultry vocalizations are increasingly recognized as non-invasive biomarkers for monitoring animal health, stress, and welfare. This thesis investigates two complementary approaches in machine learning to decode these vocal cues and push towards automated vocal monitoring in poultry. One approach uses classical acoustic feature extraction and supervised classifiers, while the other leverages NLP techniques to provide semantic analysis on vocal transcriptions. The first pipeline collected several datasets of bird vocalizations for health classification, behavioral categorization, and stress response. Audio signals were meticulously preprocessed to extract mel-frequency cepstral coefficients (MFCCs), spectral contrast, and zero-crossing rates, summarizing vocal patterns. Several machine learning models including Random Forest, Gradient Boosting, CatBoost, multi-layer perceptrons, TabNet, and LSTMs were trained and optimized via stratified cross-validation. Evaluation of their performances was based on F1-scores, Matthews Correlation Coefficient (MCC), Cohen's Kappa, and confusion matrices.

The second pipeline presents a novel multi-stage approach whereby raw poultry vocalizations are transcribed by an automatic speech-recognition system using Wav2Vec2.0 and then undergo linguistic and sentiment analysis through a BERT-based classifier. Interpretive analyses for phonetic shifts under stress and disease included token-level analyses, n-gram distributions, vowel-consonant ratios, and high-resolution word cloud plots. Parallel preprocessing using Python's ThreadPoolExecutor sped up the processing to allow for scalability when handling large amounts of audio. Results demonstrated that classical acoustic models achieved robust generalization, with ensemble methods outperforming single learners across multiple welfare contexts. At the same time, the NLP pipeline showed subtle variations of sentiment scores and linguistic complexity due to stress and health perturbations, hinting at the feasibility to transfer cutting-edge human speech models onto the semantic domain of animal vocalizations.

This work underlines the synergy between classical acoustic features and NLP-induced semantic pipelines toward a comprehensive non-invasive welfare assessment. These techniques demonstrate how AI transforms complex vocal behavior into simple indicators of health and welfare. This study lays the groundwork for the creation of an automated system for continuously monitoring the vocalizations of the birds right on the farm without disturbing them. This offers a way to detect problems early, reduce labor-intensive checks, and ensure healthier flocks with less stress. With time, this technology can support more sustainable and ethical farming by equipping farmers with dependable and real-time insights into bird welfare. This would imply better compliance with welfare standards, increased productivity, and enhanced consumer confidence toward animal rearing. In this context, the thesis presents practical insights toward the potential future for the advancement of poultry farming with the aid of an adaptive intelligent monitoring system.

List of Abbreviations Used

- AI: Artificial Intelligence
- ASR: Automatic Speech Recognition
- BERT: Bidirectional Encoder Representations from Transformers
- CASE: Clustering and Analysis of Sound Events
- CNN: Convolutional Neural Network
- CRNN: Convolutional Recurrent Neural Network
- CUDA: Compute Unified Device Architecture
- DBSCAN: Density-Based Spatial Clustering of Applications with Noise
- DSSS: Deep Supervised Source Separation
- FFT: Fast Fourier Transform
- GDPR: General Data Protection Regulation
- GRU: Gated Recurrent Unit
- GPU: Graphics Processing Unit
- HMM: Hidden Markov Model
- IOT: Internet of Things
- JSON: JavaScript Object Notation
- k-NN: k-Nearest Neighbors
- LLM: Large Language Model
- LSTM: Long Short-Term Memory
- MCC: Matthews Correlation Coefficient
- MFCC: Mel-Frequency Cepstral Coefficients
- ML: Machine Learning
- NLP: Natural Language Processing
- PAM: Passive Acoustic Monitoring
- PaSST: Patchout Spectrogram Transformer
- PANN: Pretrained Audio Neural Networks
- RF: Random Forest
- RNN: Recurrent Neural Network
- SNR: Signal-to-Noise Ratio
- SSL: Self-Supervised Learning
- STFT: Short-Time Fourier Transform
- SVM: Support Vector Machine
- TabNet: Attentive Interpretable Tabular Learning Network
- TinyML: Machine Learning on Tiny Devices
- UMAP: Uniform Manifold Approximation and Projection
- UUID: Universally Unique Identifier
- ZCR: Zero Crossing Rate

Acknowledgments

First and foremost, I would like to thank and seek the Blessings of God almighty for giving me the strength and guidance to undertake this thesis journey.

I would like to thank my extended family back home for unconditional support, love and encouragement without which I could have not come here. As someone from a joint family, being away from them for years has been extremely challenging. I could not have coped without their enduring love and support, for which I extend my humble thanks and deepest respect.

I am grateful and blessed to have Dr. Suresh Neethirajan as my Supervisor in this journey. I am especially thankful to him for his unwavering trust, abled guidance that helped me shape this thesis and my career as well.

Thanks to my thesis committee members, Dr. Bruce Rathgeber and Dr. Hassan Sajjad for taking the time to read my thesis. Thanks to Dr. Yujie Tang for graciously agreeing to chair my thesis defence.

Special thanks to my friends for being a part of this journey, for helping me, and for making it a memorable one.

Also, Thanks to Dalhousie University Libraries for providing the necessary resources, access to papers used in this thesis. A special thanks to our Research group – Mooanalytica for providing me with GPUs and my lab mates for their help.

I would like to thank my Undergraduate Supervisor Dr. Chandrakala for her guidance and encouragement which channeled me to pursue this field.

I would also like to thank Egg Farmers of Canada (EFC), Who supported this Research through multiple means.

Last, but not least, I would like to express my gratitude and fond remembrance of Shrimati M. S. Subbulakshmi, whose songs provided immense comfort and strength during this journey, especially whenever I felt down or exhausted.

Chapter 1 - Introduction

1.1 Background on Poultry Welfare Monitoring

Poultry Industry has been one of the most important backbones of global food security and also the critical source of protein worldwide [1]. Poultry animals, especially hens have been domesticated since ancient times in Asia and spread throughout the world from there [2]. Agriculture still remains in the spotlight of many countries' economies with poultry farming being an integral part of it [3]. Yet the health and welfare of poultry should not be viewed merely as a production metric, but as a linchpin to ethical husbandry and ecological sustainability. Understanding poultry well-being can amplify both economic returns and the ethical foundations of farming, aligning market objectives with compassionate stewardship.

For a long time, farmers have relied heavily on traditional methods to assess and identify the welfare and health of their animals. These often include Visual inspections, behavioral observations, swabs, blood samples or similar invasive methods from the animals. These methods are labour intensive, expensive, prone to human errors and bias and time consuming. They also sometimes fail to identify the subtle changes in the behaviour of the birds which are under stress. There is a growing concern around the world and the poultry industry that these methods are not sufficient in a large commercial farm setting as timely intervention are really important in such scenarios which in turn affects the productivity as a whole.

1.2 Vocalizations as a Non-Invasive Modality

Chickens are highly social animals. They tend to communicate a lot using their voices, body postures and actions. Poultry Vocalizations refer to the different types of sounds made by chickens and other poultry species in order to communicate with the other animals and their environment. They can be used for various purposes such as expressing their hunger, stress, pain, happiness etc. It can also be a warning or alarming others in case of threats and discovery of food as well. They reflect the intricate internal states and environmental contingencies [4]. Also, these signals extend beyond just simple communication. They serve as acoustic window to the emotional and physiological conditions [5]. For example, Chickens under stress produce a continuous, high frequency call which can recognized as a distress call. Beyond an individual level, these vocalizations can also be used to monitor a broader ecological health. Vocal complexity often represents a robust, biodiverse system. On the other hand, a restricted vocal range indicate an ecological stress or a habitat simplification. Chickens are a classic example of the complex interplay among the species in a farm ecosystem. As mentioned before, when a chicken raises an alarm call, it can sometimes influence the responses of the species that are existing with it in the farm. This strong interconnection is in sync with the acoustic hypothesis. The analysis of how chickens express their different type of calls, we can gain insights into the relationship between Prey – predator, interspecies communication and more importantly, the social dynamics of a farm ecosystem. Thus, vocalizations function as a non-invasive technique that can be used for welfare indication and also as a bioacoustic sensor for ecosystem integrity. Extending this welfare concern, the reduction in stress and an improved wellbeing can also help chickens to involve more in their natural behaviours such as foraging – this increases their ecological role. For e.g., in terms of pest control when they are comfortable, the need for chemical pesticides reduces and leads to a more sustainable agro Ecosystem. On the other hand, a shift in vocal pattern can also warn us in case of any environmental changes or degradation. This can range from extreme temperatures, poor air quality, or anthropogenic disturbances – that affects a variety of life, not just chickens. In this regard, Poultry vocalizations act like environmental feedback loops, where local stress responses may indicate a larger ecosystem imbalance.

1.3 Emergence of AI and Acoustic Sensing

The integration of artificial intelligence (AI) into livestock monitoring is reshaping the landscape of animal welfare, behavior analysis, and environmental control. In recent years, sensor-based monitoring systems have been at the forefront of research in the field of animal welfare advancement, health diagnostics, and behavioral investigation. Among all possible sensing modalities, acoustic monitoring has become a very powerful tool due to its non-invasive means of capturing biologically meaningful signals that reflect the internal states of animals. Vocalizing, jaw movements, and respiratory sounds collected through microphones or acoustic sensors can provide a continuous and real-time view of an animal's physiological and emotional states. Bioacoustic analyses have proved their potential across different species. Using acoustic sensors for jaw movement detection, for example, has been shown to accurately identify grazing and ruminating behaviors of beef cattle, providing a very fine-scale assessment method for welfare analysis in open environments [6]. Identical to the use of facial detection methods in cattle—such as the Cow Face Detection Network (CFDN)—these create individualized monitoring by integrating visual sensors into deep learning, thus advancing precision livestock management [7]. Beyond cows, these systems for deep learning-based acoustic classification have also been employed to vocalizations from elephants, allowing researchers to interpret their communication patterns based on their behaviors through spectrogram-based CNN architectures [8]. An automated voice analysis in dogs also has the capacity to identify the vocal biomarkers of emotional arousal, stress, and breed-specific characteristics, which are useful for veterinary diagnostics as well as welfare assessment [9]. In poultry, vocalizations are tightly coupled with stress, disease, hunger, and social interactions, making chickens an ideal candidate species for acoustic sensing. Microphones can passively sample an entire flock rather than disrupting their natural behavior while visual equipment might only capture a few birds at a time.

1.4 Problem statement: The need for both robust classification and semantic understanding

While stating these advantages of poultry vocalizations, we need to understand that there exists a significant gap in this research. To be precise, most of the research work done so far has been focusing on the acoustic classification of the poultry audio and treating them as black box algorithms or models. The semantic interpretability or the biological validity has been seldom considered. Moreover, Standardized datasets are not available like other tasks or scenarios. And, environmental noise robustness, lack of generalizability across breeds and contexts have made people reluctant to adopt this in real world situations. This situation is further compounded by the absence of comprehensive and critical reviews which explain the existing methodologies, identify the gaps and also one that explains the opportunities for establishing explainable, deployable acoustic sensing systems.

Therefore, there is a critical need for establishing a framework that ties the classical, statistical and semantic modeling approaches together. This enables a welfare monitoring system that is accurate and also goes beyond – integrating interpretability, biological meanings and practical deplorability.

1.5 Objectives of the Thesis

This thesis addresses the above challenges using the below three Objectives.

Objective 1: To conduct a comprehensive and critical review of the present state of poultry vocalization analysis with regards to welfare monitoring highlighting the key methodologies, constraints, and opportunities present for building explainable, deployable acoustic sensing systems.

Objective 2: To develop a framework using an integrated analytical framework that combines signal-level statistical analysis with classical machine learning and deep learning classifiers to interpret chicken vocalizations in a welfare assessment context.

Objective 3: To explore semantic and emotional decoding of poultry calls by leveraging state-of-the-art Natural Language Processing (NLP) and transformer-based models to translate bioacoustic data into meaningful insights.

1.6 Scope of the Thesis

The thesis develops on three synergistic approaches:

A critical and systematic review of the current state of research in poultry vocalization for welfare monitoring, reviewing trends, the robustness and applicability of methods, and unresolved challenges and directions for the future.

A statistical feature-based framework exploiting handcrafted descriptors and ensemble/deep learning models to classify health, stress, and behavioral situations.

An NLP-based semantic pipeline integrating wav2vec2 with BERT to determine semantic intention in vocalizations and perceive emotional cues for giving a more context-based insight into poultry welfare.

The datasets include experimentally gathered audio samples of health-related calls, stress responses, and social vocal types subjected to standard preprocessing and ethically principled decisions for data collection options.

1.7 Structure of the Thesis

- Chapter 2 provides a comprehensive and critical review of poultry vocalizations methodologies ranging from classical, deep learning approaches, semantic approaches along with Edge and IOT systems.
- Chapter 3 presents the methodology used for signal acquisition, preprocessing, and statistical classification.
- Chapter 4 describes the semantic modeling framework using NLP and transformer architectures.
- Chapter 5 reports and discusses the experimental results across both pipelines and also explores the possibility of a hybrid framework.
- Chapter 6 concludes with key insights, limitations, and directions for future research in explainable, AI-driven poultry welfare monitoring.

Chapter 2 - Comprehensive and Critical Review of Current State of Research in Poultry Vocalization

This section 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. The review entails a systematic search approach [10] as seen in Figure 2.1 through IEEE Xplore, PubMed, Sco-pus, 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").

A total of approximately 150 research works were analyzed, and out of these, 124 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. The taxonomy of our approach can be deduced from Figure 2.2. 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.

Figure 2.1 Systematic review pipeline outlining database search, screening, full-text evaluation for on-farm AI acoustic studies, and thematic synthesis

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. Several publicly available datasets—such as chick stress vocalizations [11], laying hen audio [12], and raw waveform recordings [13]—have enabled reproducible benchmarking and model comparisons.

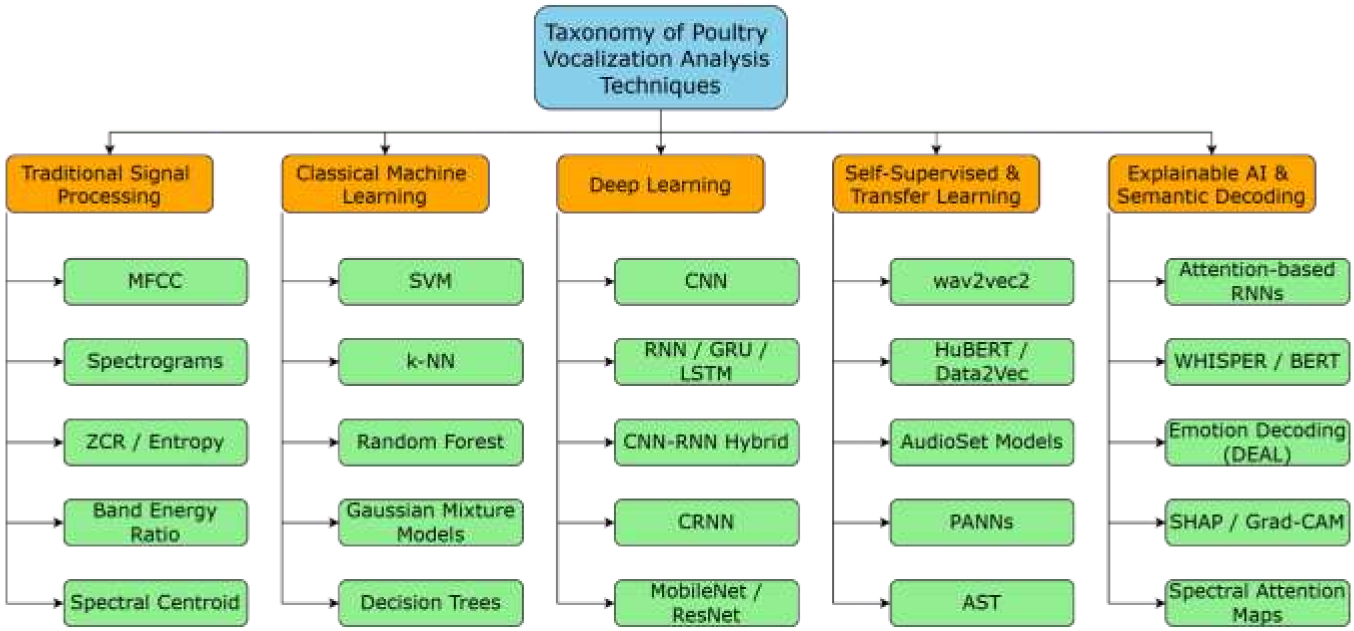

Figure 2.2 Taxonomy of poultry vocalization analysis methods across five categories

2.1 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 and have been elaborated in table 2.1. 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. [14], Pereira et al. [15], Jung et al. [16], and Thomas et al. [17] 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 [18] through certain steps of audio signal analysis that included - pre-emphasis, windowing, FFT, and DCT; compared multiple MFCC-Hybrid configurations.

Davis and Mermelstein [19] 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 [20] 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 [21] 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.

The input signals for convolutional networks also often employ spectrograms, especially log-Mel spectrograms. The work of Zhong et al. [22], Henri and Mungloo-Dilmohamud [23], Romero-Mujalli et al. [24], Thomas et al. [25], Mao et al. [26], Mangalam et al. [27], Li et al. [28], and Neethirajan [29] 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. [25] and Sainburg et al. [30]. Spectral entropy is gaining ground as a possible indicator-a-feature for distress.

Herborn et al. [5] 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. [31] had fast-induced stress in newly hatched broilers using Butterworth filtered signals and centroid spectral parameters. Tao et al. [32], 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. [33] and Soster et al. [11]. Comprehensive augmentations like frequency masking and noise injection were incorporated as stated by Mao et al. [26]. Besides feature transformation, automated segmentation tools have proven efficient, similar to the benchmark ones in Terasaka et al. [34] and Michaud et al. [12]. Merino Recalde [35] 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. [30], which sidestep traditional hand-crafted features. Swaminathan et al. [36] and Bermant et al. [33] emphasize the importance of embeddings from perusal models trained on raw audio. Du et al. [37] extracted nine temporal and spectral features based in source-filter theory to detect thermal discomfort in laying hens.

Table 2.1 Comparison of static and dynamic acoustic feature sets in animal vocalization studies.

Feature Name	Study	Model Used	Environment	Reported Accuracy	Notes
SincNet	Bravo Sanchez et al. [13]	Raw waveform classifier	Minimal preprocessing	>65% (NIPS4Bplus)	Learns directly from waveform, robust to distortions
MFCC	Umarani et al. [7]	LSTM	General (RAVDESS)	97.22%	LSTM + MFCC for emotion recognition
MFCC	Jung et al. [16]	CNN	General	91.02% (cattle), 75.78% (hens)	Lower for hens—possibly due to background noise
MFCC variants + FFT/DCT	Prabakaran & Sriuppili [18]	MFCC variants	Controlled	94.44%	Comparative setup across MFCC variations
Cochleagram	Sattar [20]	Context-aware classifier	Noisy farm	>20% higher than MFCC	Better adaptability to environmental noise
Mel-Spectrogram	Henri et al. [23]	MobileNetV2	Birdsong (natural)	84.21%	Limited context modeling
Spectral Entropy	Herborn et al. [5]	Entropy analysis	Chick stress study	Qualitative improvement	Captures emotional states during distress
Wav2vec2 Embeddings	Swaminathan et al. [36]	Fine-tuned classifier	Real-world bird data	F1 = 89%	SSL embeddings outperform handcrafted features
MFCC	Bhandekar et al. [38]	SVM	Lab	95.66%	Strong in low-noise environments

2.2 Classical and Deep Learning 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. 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 as seen in table 2.2 and 2.3 For example, Bhandekar et al. [38] 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%.

Pereira et al. [15] reported 85.61% accuracy with a Random Forest model trained on FFT-extracted features to assess the distress of chicks. Tao et al. [32] 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%. Ginovart-Panisello et al. [39] used Gaussian Naive Bayes in detecting vaccine response classified based on MFCCs and spectral centroid, with an F1-score of 80%. Du et al. [37] applied SVMs to temporal-spectral features toward the detection of thermal discomfort at a sensitivity of 95.1%.

Table 2.2 Performance of classical machine learning models in animal vocalization classification.

Authors	**Model(s)**	**Reported Accuracy**
Pereira et al. [15]	Random Forest	85.61%
Tao et al. [32]	SVM, RF, k-NN	94.16%
Du et al. [37]	SVM	Sensitivity = 95.1%
Bhandekar et al. [38]	SVM	95.66%
Ginovart-Panisello et al. [39]	Gaussian Naive Bayes	F1-score = 80%

Table 2.3 Performance of deep learning architectures for animal vocalization classification, including CNN, RNN, and attention-based models.

Authors	**Model Type**	**Reported Accuracy**
Jung et al. [16]	2D CNN	91.02% (cattle), 75.78% (hens)
Henri et al. [23]	MobileNetV2	84.21%
Romero-Mujalli et al. [24]	DeepSqueak CNN	Detection: 91%, Class: 93%
Mao et al. [26]	Light-VGG11 CNN	95%
Mangalam et al. [27]	Lightweight CNN	92.23%
Hassan et al. [45]	Conv1D + Burn Layer	98.55%
Hu et al. [47]	MFF-ScSEnet (attention)	>96%
Hu et al. [47]	MFF-ScSEnet CNN	>96%
Gupta et al. [43]	CNN-LMU	Best model
Mousse & Laleye [46]	Attention-based RNN	F1-score = 92.75%

Henri and Mungloo-Dilmohamud [23] compared MobileNetV2, InceptionV3, and ResNet50, with MobileNetV2 achieving 84.21% accuracy. Mangalam et al. [27], a lightweight custom CNN

(~300k parameters), outperformed fine-tuned VGG16. Mao et al. [26] discovered light-VGG11 with a 92.78% decrease in parameters, which retained 95% accuracy. Ginovart-Panisello et al. [39] used CNNs trained on spectrograms for detection of stress. Cuan et al. [40, 41]: CNN-based detection of Newcastle disease and avian influenza. Ginovart-Panisello et al. [31]: CNNs (ResNet) for detection of acute stress based on vocalization and thermographic data. Li et al. [28]: ResNet-50 trained on MFCC+Logfbank features for chick sex detection.

Research utilizing temporal modeling via RNNs, LSTMs, GRUs, and hybrid CNN-RNN models appear often in literature dealing with the sequential structure of vocalizations. Umarani et al. [14] and Bermant et al. [33]; Li et al. [28] and Xu and Chang [42]; Gupta et al. [43] and Jung et al. [16] are examples of GRU, CRNN, and CNN-LMU evaluations. Huang et al. [44] developed a sequence model to detect poultry feeding behavior based on vocal patterns.

2.3 Hybrid, Attention, and Benchmark Architectures

A Conv1D-based classifier with Burn Layers (noise-injection modules) was implemented by Hassan et al. [45] to enhance generalization, leading to an impressive accuracy of 98.55%. Mousse and Laleye [46] established an attention-based RNN for hens' behavior recognition and reported an F1 score of 92.75%. Hu et al. [47] proposed MFF-ScSEnet, which combines Mel-spectrogram and SincNet features with a squeeze-and-excitation mechanism and more than 96% accuracy.

Ginovart-Panisello et al. [39] and Thomas et al. [25] have performed both ablation studies and multi-objective training (classification + age estimation). Bermant et al. [33] benchmarked CNNs and RNNs across echolocation and coda recognition tasks and got over 99% accuracy. Gupta et al. [43] and Ghani et al. [48] 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. [49] 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.

2.4 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.

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 [23] who refined MobileNetV2, ResNet50 and

InceptionV3 for bird song classification, with best accuracy (84.21%) corresponding to MobileNetV2. Thomas et al. [25] transferred PANN (Pretrained Audio Neural Network) weights to a multi-objective CNN for broiler vocalization and age estimation. Mangalam et al. [27] compared a custom CNN with fine-tuned VGG16, concluding that the smaller model worked better under field conditions. Li et al. [28] showed that chick sexing tasks conceived from different architectures (ResNet-50, GRU, CRNN), based on breed and feature type, perform variably. McGinn et al. [50] obtained unsupervised feature embeddings derived from the BirdNET CNN to classify within-species vocalizations, emphasizing its strength without retraining. Ginovart-Panisello et al. [39] applied pretrained CNNs to the spectrograms of hens to induce stress response for vaccinated hens. Vaswani et al. [51] 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.

In a more foundational review concerning AI in livestock, Menezes et al. [52] 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. [53] 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. Ghani et al. [48] 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. [36] 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. [54] 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 [55] found speech-pretrained SSL models to perform at par with those trained specifically for bioacoustics, making it feasible to reuse human-centric models. Neethirajan [56] 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. [57] used Transformer-based models for long-range dependency studies in Bengalese finch songs: eight syllables appeared to be good context length. Gong et al. [58] introduced the Audio Spectrogram Transformer (AST)—a 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. Baevski et al. [59] 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., [36], [54]. Wang et al. [60] 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. [61] 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. [13] 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. [62] 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. Wav2vec2 performs better than many traditional models in poultry call detection because of its combination of contextualized audio embeddings and contrastive self-supervised training. In general, the MFCC pipeline depends on handcrafted features, but wav2vec2 learns deep representations from a raw waveform by predicting masked latent representations. In this way, the model is able to catch subtle temporal patterns and contextual variations in vocalizations and distortions that degrade standard features in a noisy farm environment. Its fine-tuning possibilities with limited labeled data also make this model apt to be used in low-resource domain problems such as poultry welfare monitoring.

Similarly, SincNet performs better over several CNN-based methods due to its ability to learn sinc-based filters that are constrained to represent meaningful frequency bands that are to be valid frequency bands. This inductive bias enables the model to extract frequency-specific features that are physiologically relevant to bird calls while reducing the parameter search space, thus enhancing generalization across small datasets. Lastly, it operates on the raw waveform directly, avoiding any possible errors introduced in transformations to the spectral domain, such as STFT or Mel-scaling, giving the classifier increased resilience to varying acoustic distortions encountered in the real world.

While models like wav2vec2 and Whisper fine-tuned for poultry vocalizations perform exceedingly well, one should observe that their original training was always conducted on human-speech corpora. The structure, phoneme inventory, and temporal dynamics of animal sounds are far from those of human speech. Consequently, although such systems can offer a generic resolution to acoustic feature extraction, the semantic alignment and acoustic priors engineered for human speech do not offer the best clues for the decoding of emotional or behavioral cues speciated to poultry. For instance, spectral bandwidth and non-verbal call structures of birds lack phonetic segmentation assumptions that human speech models rely heavily upon. Mismatches like these become sources of acoustic noise on downstream tasks, which limits zero-shot generalization to presence across unseen animal domains.

In addition to the manual transfer learning, some studies employ an active nudging from automated approaches in discovering models: Tosato et al. [63] established an optimal Xception architecture for classifying bird vocalizations by using AutoKeras that is better than MobileNetV2, ResNet50, and VGG16. Gupta et al. [43] 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.

2.5 Emotion and Stress Detection in Poultry Vocalizations

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 [64], 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. [65]. This revealed beak and comb temperature reduction and decreased call rate following stressor exposure. In similar fashion, Ginovart-Panisello et al. [31] 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. [5] 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. [66], who related these to various levels of emotional arousal as represented by loudness, frequency, and duration. Lev-Ron et al. [67] 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. [68] 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. [69], 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.

2.6 Behavior Recognition and Emotional Modeling in Poultry

Behavioral responses are mirrored in voice patterns. Zimmerman [70] first worked on the "gakel-call" in hens and established linkages with the emotion of frustration that stems from blocked behaviors. More recently, Mcgrath et al. [4] 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 study conducted by McGrath et al. [71] revealed that people could identify the chicken calls reliably associated with rewards, indicating the presence of semantic information encoded within the calls. Neethirajan [72] also studied this topic with the WHISPER model, confirming token-based patterns in chicken distranquil vocalization correlated to emotion. Abzaliev et al. [73], 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.

Schober et al. [74] 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. Neethirajan [75] 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. [76] developed the DEAL model (Deep Emotional Analysis Learning) to interpret emotional states such as hunger and fear in chickens. Ginovart-Panisello et al. [39] 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. [37] 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. [77] 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 [78] underlined emerging techniques in passive acoustic monitoring (PAM) for emotional state estimation, pointing to foundational models and threshold-free density estimation tools.

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. [44] have established a 95% accuracy rate for identifying episodes of eating behavior in chickens. Using attention-based RNNs, Laleye and Mousse [46] classified laying hen behaviors with an F1-score of 92.75%. Fontana et al. [79] found a negative correlation between broiler vocal frequency and weight, thus establishing an association between acoustic cues and physiological growth. Karatsiolis et al. [80] proposed a non-invasive farm monitoring system that uses vocal, visual, and environmental sensor data to- interpret Flock-wide psychological states. Manteuffel et al. [81] reviewed how vocal correlates—like call frequency and formant dispersion—indicate both positive and negative emotional states in multiple species of live-stock. Güntürkün [82] 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, in —cognitive load, or interest state. Galef and Laland [83] 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. [84] 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.

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) [85]. In particular, vocal measures of distress, anticipation, and contentment correspond to a kind of "Mental State" do-main—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 [86]. 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.

2.7 Disease Detection and Health Monitoring in Poultry

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. 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 [1]. 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. [40] 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. [41] 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 [42] 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 [56] used Whisper, which took chicken vocalizations and created token sequences that were sentiment-scored to identify emotional states and physiological states. Adebayo et al. [87] 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.

Du et al. [37] 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. [28] 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 [21] 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. He et al. [3] 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. [26] 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. [11] 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. [17] 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. [88] 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. Srinivasagan et al. [89] 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. Huang et al. [44] 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.

2.8 End to End Pipeline and Automation

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.

Bioacoustic software tools for automating large parts of the workflow have recently emerged and have been represented in figure 2.3. Gibb et al. [90] 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. [91] 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. [17] 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.

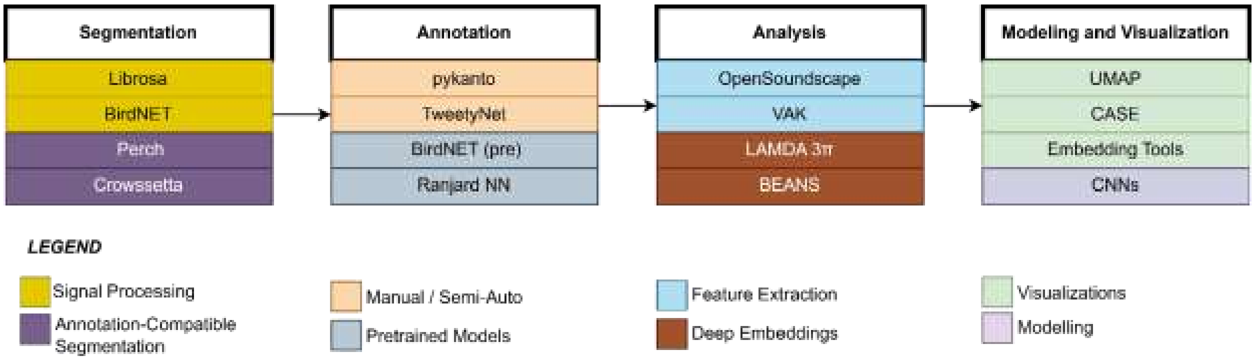

Figure 2.3 Workflow of Bioacoustic Analysis: Segmentation to Modeling using Specialized Tools.

Merino Recalde [35] 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 [92] 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. [93] 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. [94], 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.

For reliable segmentation, high-quality training datasets are essential. In this context, Terasaka et al. [34] 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. [12] 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. [95] 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 [96] 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. [97] through end-to-end training, demonstrating good generalizability across species. Lastly, Sethi et al. [98] 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. Lostanlen et al. [99] 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. [100] 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. [101] 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 [102]. Bermant et al. [33] 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. [103] 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. [104] 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.

2.9 Edge AI and TinyML: Enabling Practical, Low-Power Acoustic Monitoring in Poultry Farming

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. With edge AI, mainly through TinyML, real-time inference is performed directly on equipment deployed at farms. In this way, Srinivasagan et al. [89] 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. [88]). Using phase-coding and Gaussian classifiers such as SVM and k-NN on hardware of Raspberry Pi, Bhandekar et al. [38] designed a real-time monitoring system for analysis with synchronized video and audio tracking. Huang et al. [44] developed a module of vocal formants to detect the feeding behavior in noisy field conditions. 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 [105]. 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 [106].

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.

Studies that have addressed the effects of noise and changing environments: Mao et al. [26] 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. [27] 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. [107] 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. [80] 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. Long-term field studies conducted by Ginovart-Panisello et al. [108, 109, 110] 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. [39] 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. [31] tracked call rates and spectral features in real-time. Niu et al. [111] 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. Many studies involve optimization to reduce model size, boost energy efficiency, or simplify their architecture: Mao et al. [26] reduced the total number of parameters by 92.78% against standard VGG11. Hassan et al. [45] introduced Burn Layers (noise-injection modules) to improve generalization under deployment noise. Ginovart-Panisello et al. [39] 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.

2.10 Bridging Gaps in Poultry Bioacoustics: Data, Interpretability, and Responsible Deployment

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. [112]. 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 [78] describe low signal-to-noise ratios, class imbalance, and lack of global standardized datasets. Coutant et al. [113] 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. The question of whether models trained on one species or domain generalize to another is central to future applications. Van Merriënboer et al. [114] reviewed evaluation methods and showed how data variability and covariate shift affect degradation in generalization. Ghani et al. [48] and Gupta et al. [43] showed that transfer learning improves performance, but it still incurs a performance drop in unseen soundscapes or under polyphonic conditions. Swaminathan et al. [36] and Sarkar and Magimai-Doss [55] 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. There arises a gap especially when transfer is considered from speech-pretrained models, such as wav2vec2 and Whisper, to the domain of poultry vocalizations. These models are trained on signals that resemble structured language, which include phonemic regularities and sentence-level dependencies. Calls from poultry are short, affective, continuous, or rhythmic and lacking in segmental structures. In the absence of fine-tuning for the given task, these models may be unable to relate acoustic patterns to a meaningful biological interpretation.

Domain mismatch produces embedding shifts where from one context of species/environment, feature representations learned would be misaligned in another environment. For example, models trained on chick or hen vocalizations usually fail to generalize to duck calls because calls are species-specific and can differ in harmonic structure, call duration, and frequency modulation. Swaminathan et al. [36] and Ghani et al. [48] observed that the fine-tuned wav2vec2 and PaSST models performed well and produced high accuracy within each specific dataset, yet embed drift occurred since they performed poorly or with reduced accuracy when tested on datasets of different species or recording setups. Ginovart-Panisello et al. [31] also reported failures in cross-breed generalization when training stress detection models on broiler vocalizations and applying them to laying hens.

Although many have succeeded in high classification rates, the number of works which deal with interpretability of vocal signals is less. Neethirajan [72] and Cai et al. [76] 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 efforts, according to Stowell [115]. While many advanced models, including CNNs, RNNs, and Transformers, manage to achieve high levels of classification accuracy, they are less interpretable, particularly so in realms requiring trust and transparency, such as animal welfare monitoring. The nature of the deep learning paradigm constitutes the so-called "black-box" problem.

For instance, wav2vec2 and Whisper models achieve highly accurate classifications yet offer little insight into which vocal features or temporal patterns it has relied on. Grad-CAM, SHAP, LIME, among others, are seldom used in bioacoustics; even when they are, the focus tends to be on spectrogram-level saliency as opposed to some biologically meaningful acoustic markers. Bolhuis et al. [116] reject claims of syntactic structure in bird vocalizations, stating that animal communication lacks true combinatorial semantics. Berthet et al. [103] supports the importation of linguistic theories into animal communication (i.e. syntax, pragmatics) and argues that such models should respect certain ethological constraints. Jarvis [117] brought together many lines of research in vocal learning to suggest that animals might share features of language.

Takeshita and Rzepka [118] identified numerous NLP datasets and models as embedding speciesism, thus warranting the need for the fair representation of nonhuman vocalizations in research and applications. According to Zimmerman [70], Mcgrath [4], Manteuffel et al. [81], and Marino [119], there is a pressing need for further behavioral and emotional interpretations of poultry vocalizations. 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. Techniques for noise cancellation can be explored through spectral subtraction, Wiener filtering, and neural-based speech enhancement [120]. 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) [121]. 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 [122]. 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 [123]. In order to achieve deployment transparency, XAI solutions can be used. For instance, Grad-CAM or LIME visualization techniques [124] 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 [125] that will allow farmers and technicians to actively engage in annotation, validation, and deployment, cultivating long-lasting adoption and trust toward the technology.

The studies thus far speak to recurring challenges-from lack of availability of domain-specific data and signal variability to the interpretability bottlenecks confronted by black-box architectures-that call for more granular, multi-faceted analytical approaches. In answer to these challenges, this thesis describes a dual-pathway framework aimed at the systematic dissection of poultry vocalizations from statistical and semantic standpoints. The following chapter (Methodology I) embarks on this journey by comprehensively providing details from acoustic feature acquisition, preprocessing, and extraction processes alongside the classical and deep learning classifiers used to assign welfare-related conditions to vocal signatures, thereby setting the stage for further explorations of temporal dynamics and semantics toward the holistic decoding of vocal indices of poultry health and stress

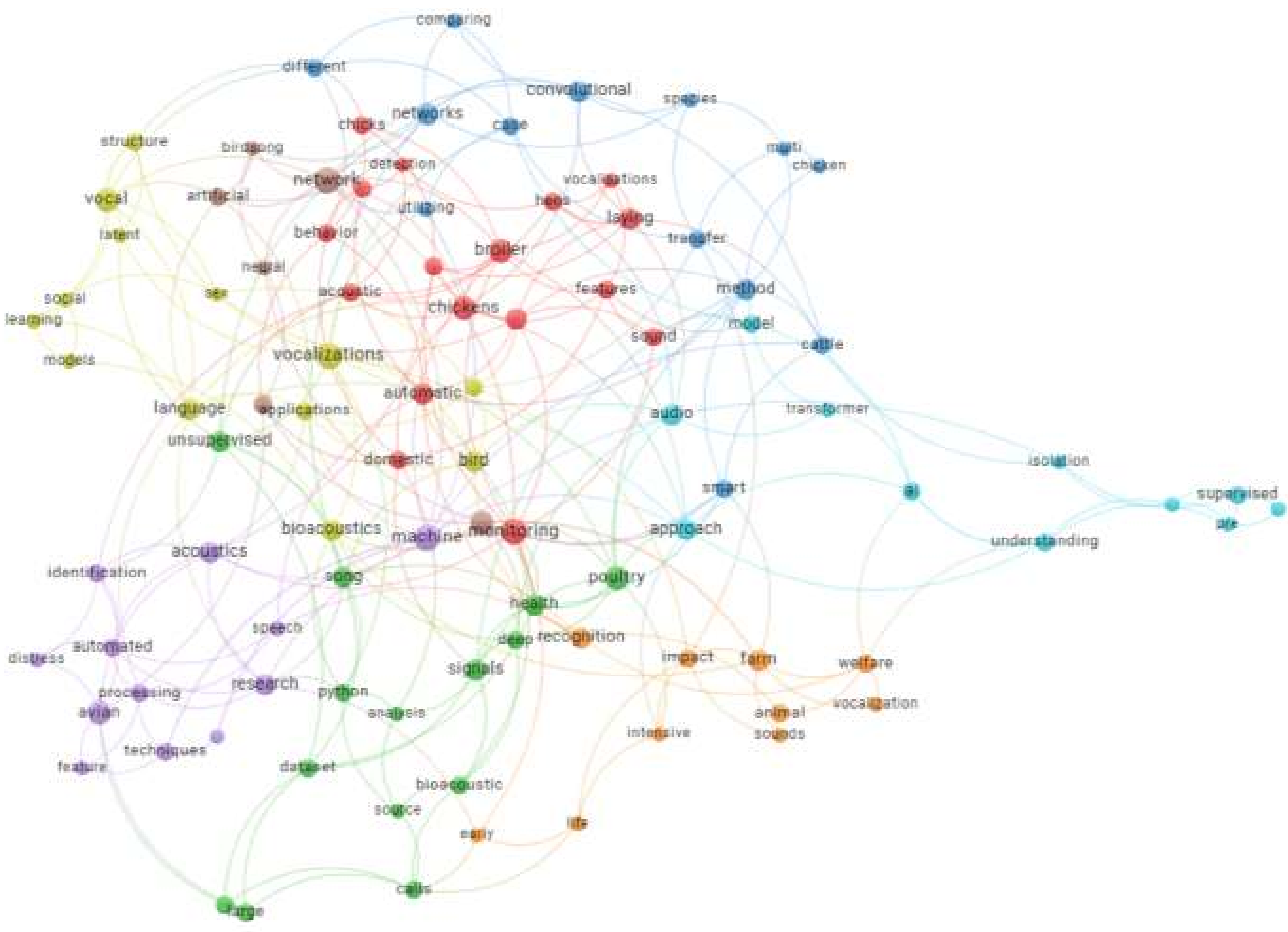

Figure 2.4 Keyword co-occurrence network showing thematic clusters in livestock vocalization research. Node size indicates keyword frequency, while colors represent distinct research themes such as poultry monitoring, acoustic analysis, and deep learning approaches.

Chapter 3 - Methodology I: Statistical and Temporal Approaches for Poultry Vocalization Analysis

Figure 3.1 Acoustic sensing pipeline for poultry monitoring, including signal acquisition, preprocessing, feature extraction, statistical analysis, and machine learning-based classification of health, behavior, and stress states.

3.1 Data Sources

Three public datasets of poultry vocalizations were selected for this study to cover a diverse range of welfare-related contexts, including health classification, behavioral categorization, and stress response:

3.1.1 Dataset 1:

The experimental dataset (10.17632/zp4nf2dxbh) used in this study was collected from a controlled poultry research facility at Bowen University, Nigeria, where 100 day-old broiler chicks were split into treated and untreated groups and raised under identical monitored conditions. Audio data was collected over a 65-day period, thus covering vocalizations from market-age broilers (35–42 days and beyond). While the exact commercial line (e.g., Ross 308 or Cobb 500) was not specified in the dataset documentation, the study was explicitly conducted on broiler chickens, suggesting a fast-growing breed of standard commercial strain. The dataset also did not report the sex of the birds; however, since broilers are often reared in mixed-sex flocks, it is likely that the recordings represent sex-combined data, which aligns with typical commercial farming practices. Future work should aim to incorporate fine-grained metadata (e.g., genotype, sex, weight, and growth rate) to explore potential breed- or sex-specific acoustic signatures. For respiratory illnesses, the first group received treatment, whereas the second group did not. After 30 days, the untreated group started to sound sick with respiratory issues. This information was also noted as being unhealthy. The main aim of the dataset is to distinguish between healthy and unhealthy birds. In total, the dataset used has 52 labeled audio samples classified into two classes (28 healthy and 24 unhealthy) each denoting general health and disease situations such as respiratory infection. Recording took place in low-noise laboratory environments, using directional microphones sampling at 96 kHz. The average duration of each clip is approximately 5–60 s. All files were converted to mono and standardized in sample rate prior to feature extraction. Spectral subtraction and silence trimming were applied to signals so that preprocessing improved signal clarity.

3.1.2 Dataset 2:

The ChickenLanguageDataset (https://github.com/zebular13/ChickenLanguageDataset) consists of a very huge collection of audio samples capturing chicken vocalizations in different behavioral and social contexts. This dataset contains labeled samples for vocal types like tidbitting, do_you_have_food, where_is_everyone, hungry, and eating. The recordings were taken using electret condenser microphones installed in a pen with little human interference, in natural ambient noise and lighting. The audio files were sampled at 48 kHz and last from 2 to 7 s. This dataset was intended primarily for tasks involving behavioral classification. Noise gating and dynamic range normalizations were carried out before modeling to remove external noise, and the final class-wise distribution was checked for imbalance before under-represented classes were kept after synthetic balancing using stratified sampling during cross-validation.

3.1.3 Dataset 3:

This dataset contains the vocal responses of chickens exposed to both visual stressors (the opening of umbrellas) and auditory stressors (dog barking). The dataset used is from CARUS Animal Experimental Facility, Wageningen University (https://zenodo.org/records/10433023). A total of fifty-two Super Nick laying hens were systematically distributed among three distinct cages within a stable-sized experimental room configured to closely replicate commercial poultry production conditions. To further mitigate auditory contamination, the control cage featured robust acoustic shielding composed of an 18 mm thick plywood panel supplemented by three layers of 6 mm high-density corrugated cardboard, achieving a combined barrier thickness of 36 mm. The facility adhered strictly to CARUS guidelines, controlling temperature, humidity, and air quality via distributed sensors. Longitudinal recordings were taken in different weeks to track adjustment or sensitization over time to that stress. The dataset contains more than 200 audio segments labeled according to the type of stressor and the week of recording. The placing of microphones was strategic enough to catch the vocalization responses while avoiding too much interference from equipment. During preprocessing, partial mitigation of background noise such as barn fans was achieved through high-pass filtering and spectral noise reduction. On average each clip lasts around 4–6 s. This dataset enabled time-series trend analysis of stress-induced vocal features. All animal vocalization recordings used in this study complied with the respective institutional animal care and use guidelines. No experimental harm or behavioral disruption was introduced.

3.2 Acoustic Sensing and Signal Acquisition

Sensors are key instruments in advancing precision livestock systems, making it possible to non-invasively inspect animal health and welfare in the real world. Advantageous to poultry production, acoustic sensing provides a medium for the utmost purpose, since birds' vocalizations house rich, quantifiable cues about physiological state, emotional responses, or behavioral changes. In contrast to visual monitoring or tagging with wearables, microphones are inexpensive, scalable, and can continuously monitor the behavior of huge flocks without them knowing or causing stress.

In this study, acoustic sensing formed the foundation of a data-driven pipeline (Figure 3.1) for evaluating poultry welfare. The audio signals were recorded in controlled environments using external microphones set for high-fidelity recording. These vocalizations, taking place under certain behaviors or physiological conditions (feeding, distress, health deterioration), are natural outputs of internal states and, therefore, passive signals that biologically give information about the animals' states

3.3 Preprocessing and Feature Extraction

There are four major stages in the pipeline for studying the acoustic data:

Signal Acquisition: High-fidelity audio was recorded at poultry farms or research environments using condenser microphones set at strategic locations for good pickup of vocal outputs

Preprocessing: The signals were cleaned and segmented using various techniques like spectral subtraction and windowed framing to eliminate the background noise for feature quality enhancement.

Feature Extraction: Acoustic features were computed from the recordings, e.g., mel-frequency cepstral coefficients (MFCCs), spectral contrast, zero-crossing rate (ZCR), and several other temporal and spectral descriptors. These features statistically represent parameters of sounds that pertain to exertion, tonality, rhythm, and structural properties of the vocalizations.

Classification: Vocalization classification into known classes-such as healthy or unhealthy, behavioral types, and stress responses-was performed using supervised machine learning. The models combined conventional ensemble methods with neural and deep learning methods.

Features like MFCCs, spectral contrast (example plot in figure 3.2), and zero-crossing rate provide a compact statistical summarization of the vocal signals' frequency spectrum, energy distribution, and their temporal dynamics. As an example, while MFCCs are useful in modeling the perception auditory scale, are sensitive to changes in the volume of the vocal tract, and spectral contrasts capture the harmonic structure, ZCR captures the irregularity of the signal. Respiratory health, arousal, and stress have biological inflence of these features. These features mapped out the raw audio so that they can be understood by transforming them into an intelligible space where machine learning is able to work on them and relate unique vocal patterns to physiological and behavioral states in poultry.

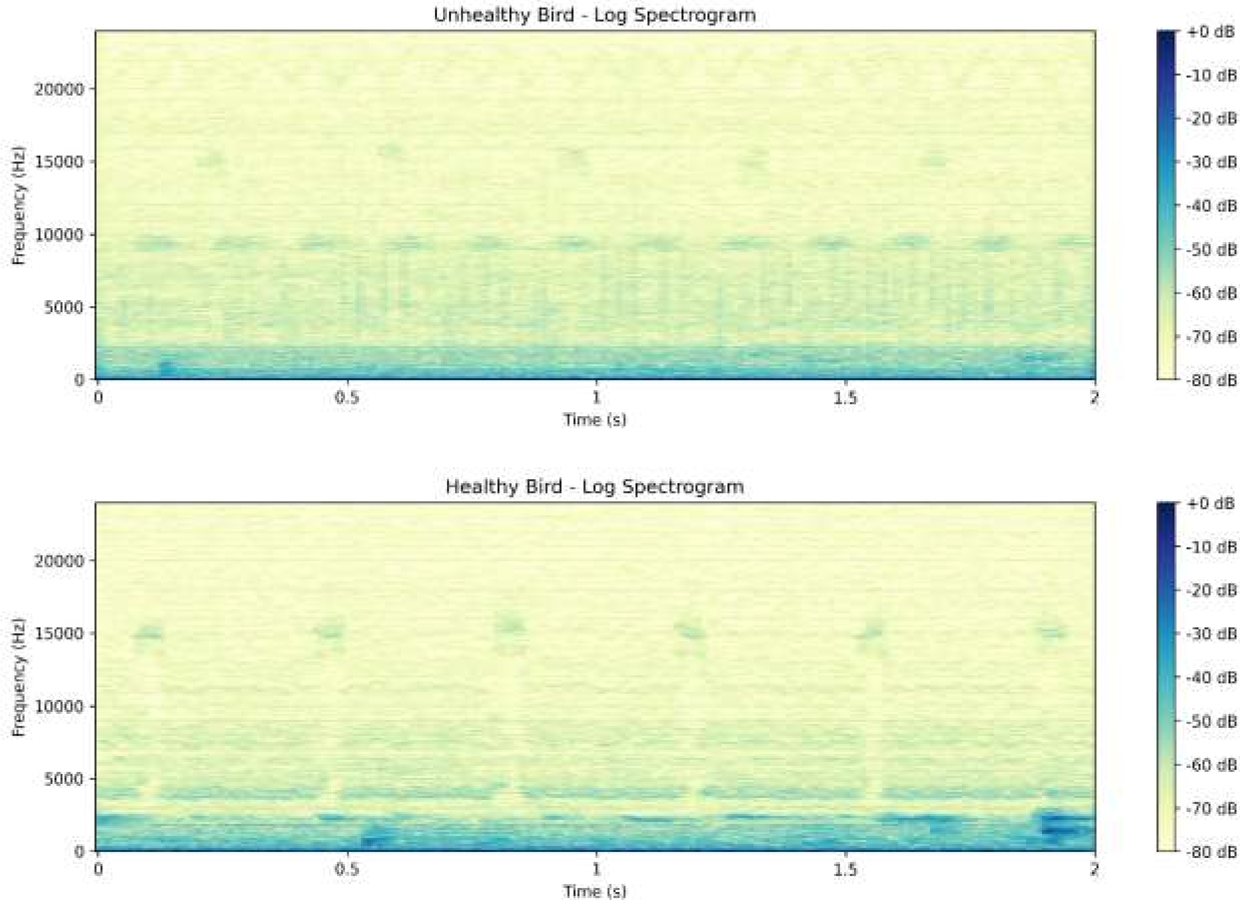

Figure 3.2 Log spectrograms for healthy and unhealthy birds (0–2s). Unhealthy vocalizations show reduced spectral richness and elevated low-frequency energy, indicating vocal strain.

3.4 Temporal LSTM

In addition to static classification models, we included an **LSTM network** for the task of stress detection on Dataset 3. This dataset consisted of recordings across weeks and was therefore longitudinal. Unlike health or behavior classifications, this particular dataset captures temporal evolution over the vocal responses to stress.

LSTMs are used since they are very suitable for retaining temporal dependencies and modeling sequential patterns of input data over the long run, settling these issues very well for time-series input data. To study the dynamic changes in vocal features over time such as habituation to stress or delayed responses, which are difficult for models lacking temporal memory to capture, LSTM would be most appropriate. Hence, aligned with both the nature of the biology and that of the experiment, we were able to identify stressor-specific trajectories within vocalization patterns using LSTM.

3.5 Cross-Validation Strategy

Every model was evaluated using **stratified five-fold cross-validation** to ensure a balanced presence of classes across the folds and to avoid any biases on learning arising due to uneven share of class distributions. During every fold, 80% of the data was used to train the models while the remaining 20% served validation.

This procedure was carried out five times, and therefore, the final performance metrics were computed as an average across the folds. Thus, the whole exercise would provide robustness against variance in the training-test splits, and thus more reliable estimates of model generalization.

3.6 Evaluation Metrics

To give a really broad view of classification performance, a collection of standard supervised learning metrics was chosen that took into account the fact of bias and misleading prediction errors that occur in biological data.

- Accuracy indicates how many instances were correctly classified in overall proportion, but it may become misleading in contexts of imbalanced datasets where one class dominates others.
- Hence, we report:
    - Precision (TP / predicted positives),
    - Recall (TP / actual positives),
    - F1-score (the harmonic mean of precision and recall), since they give weight to false positives and false negatives.
- We further include:
    - Cohen's Kappa, which measures agreement between predicted and true labels while accounting for chance,
    - Matthews Correlation Coefficient (MCC), which provides a balanced measure of binary classification performance even under skewed class distributions.

Such metrics are highly recommended for biomedical and bio-acoustic classification tasks due to their resistance against imbalanced data and interpretability.

3.7 Hyperparameter Tuning

Hyperparameter tuning (Table 3.1) was carried out by RandomizedSearchCV, each model with different parameter grid specifications. The optimal configuration for every model was chosen based on validation accuracy during three-fold cross-validation, followed by evaluation under five-fold stratified cross-validation. We explored grid search and Bayesian optimization with Optuna but maintained a randomized search for tuning hyperparameters. This made it easier to keep things consistent and compare the different classifiers fairly.

Table 3.1. Hyperparameters Tuned for Each Classifier

Classifier	Hyperparameters Tuned
Random Forest	n_estimators ∈ [100, 200], max_depth ∈ [None, 10, 20]
Extra Trees	n_estimators ∈ [100, 200], max_depth ∈ [None, 10, 20]
Gradient Boosting	learning_rate ∈ [0.01, 0.1], n_estimators ∈ [100, 200], max_depth ∈ [3, 5, 7]
Adaboost	learning_rate ∈ [0.01, 0.1], n_estimators ∈ [100, 200]
Catboost	iterations ∈ [100, 200], learning_rate ∈ [0.01, 0.1], depth ∈ [4, 6, 8]
Histgradientboosting	learning_rate ∈ [0.01, 0.1], max_depth ∈ [3, 5, 7], l2_regularization ∈ [0.1, 1]
Mlpclassifier	hidden_layer_sizes ∈ [(50, 50), (100, 50)], activation ∈ [‘relu’, ‘tanh’], alpha ∈ [0.0001, 0.001]
Tabnet	n_d, n_a ∈ [8, 16, 24], n_steps ∈ [3, 5, 7]

Chapter 4 - Methodology II: Semantic and Emotional Decoding of Poultry Vocalizations

4.1 Preprocessing Signal Segmentation, and Parallelized Execution

Preprocessing involves converting and segmenting raw poultry vocalization recordings into formats amenable to speech-to-text transcription and linguistic analysis. This phase is pivotal in changing continuous audio streams into analyzable semantic units. The raw audio files were recorded in either .mp3 or .wav and stored in structured directories pertaining to a particular condition, such as, PoststressTreatment, PrestressTreatment, or Disease groups (i.e., healthy, unhealthy, and noise). These files were batch-processed in parallel by leveraging Python's ThreadPoolExecutor; this approach permitted processing multiple audio files at once. Parallelization significantly blocked the bottlenecks caused by I/O operations and thus sped up the preprocessing, in particular when working with large audio datasets constituting extended recordings from the farm. The .mp3 files were converted into .wav format first with the AudioSegment class provided by the pydub library. While converting, the sampling rate was fixed to 16 kHz, and downmixing to a single channel (mono) was performed. These settings need to be consistent as per the input requirements of the downstream Wav2Vec2 model. Partitioning audio files is a requirement to hinder the exceeding of length-out-of-limitations for automatic speech recognition models and to preserve transcription availability in case ambient noises or overlapping vocalizations occur.

4.2 Transcription via Wav2Vec2.0 for Textual Corpus Creation

Each audio segment was then passed to a transcription pipeline using Wav2Vec2.0, employing the pretrained facebook/wav2vec2-base-960h model accessed through the Transformers library of Hugging Face. Wav2Vec2.0 is a self-supervised speech representation model trained over 960 hours of the Librispeech audio database.

Unlike traditional ASR systems relying on spectrogram extraction and separate acoustic models, Wav2Vec2 ingests raw waveform inputs directly, through a stack of temporal convolution layers followed by transformer encoders. Inside the network, it learns embeddings of audio contexts by contrasting masked segments against negative examples-a technique that allows the model to acquire rich phonetic and semantic representations already before fine-tuning. In this implementation, we loaded each audio waveform with the librosa library and resampled it to 16 kHz to fit the input resolution expectations. The waveform was tokenized by the Wav2Vec2 processor, which returned input_values tensors that were transferred to the CUDA device to perform inference using GPU acceleration. The model then outputs a tensor of logits through a vocabulary of characters, and greedy decoding (argmax) was used to get the most likely transcription. The decoded text of each segment was concatenated to the running transcription string. Finally, the segment files were deleted to free space and maintain cleanliness. The transcriptions from each directory (e.g., Poststress, Prestress, etc.) were collected and stored in

JSON files or dictionaries for further NLP-based sentiment analysis and classification. These structured textual data formed the semantic basis on which all subsequent linguistic and classification work was based.

4.3 Semantic Sentiment using BERT and Phonetic Composition Analysis

Moving on to the emotional content of the transcribed content, we used a pretrained multilingual BERT model (nlptown/bert-base-multilingual-uncased-sentiment). After tokenization, padding, and truncation to a fixed length of 512 tokens, the transcripts were passed through BERT on a CUDA-enabled GPU. The output logits were transformed via softmax to yield normalized scores across the three sentiment categories: positive, neutral, and negative. The sentiment proportions were then accumulated for all transcriptions in a condition and depicted using high-resolution pie charts so that the emotive valence could be directly compared among different health- or stress-related circumstances.

NLP based feature-extraction was used to study the types of vocabulary employed, phonetic orientation, and lexical complexity. Transcriptions were tokenized into individual words using NLTK's `word_tokenize` function, which breaks down string texts into manageable tokens (eg. "The chick is calling" -> ["The", "chick", "is", "calling"]). Using those tokens, one could generate frequency distributions for unigrams (single words) and bigrams (two-word sequences) using `Counter` from Python. Additional counts were made for individual characters and vowels (a, e, i, o, u), so as to analyze phonetic emphasis and shifts tied to welfare conditions.

Frequency plots of unigrams, bigrams, and characters give a numerical depiction of any change in vocal repertoire complexity with either health or stress. Bar plots were drawn for the top 20 unigrams and bigrams, whereas the upper 25 characters were presented to highlight repeating phonetic patterns along the chicken vocal lexicon. Separate vowel frequency plots amplified the biological interpretation that vowel-heavy utterances correspond to stressed and diseased birds, whereas fewer vowels correspond to reduced vocal complexity. To complement the statistical measurements, word clouds were generated with the WordCloud library; these depicted the most frequently appearing words scaled by their relative size, offering an intuitive view of prominent vocal expressions present in each experimental condition.

4.4 BERT-Based Supervised Classification on Acoustic Embeddings

A separate supervised learning setup (table 4.1) was also designed to directly classify the acoustic states from the inherently structured embeddings of MFCCs and spectral features. These embeddings were then coupled with categorical labels combining stress treatment type and recording week (WeekX_StressY) to study stressor-specific temporal dynamics. The custom PyTorch architecture was built from bert-base-uncased and was made to take direct feature embeddings via inputs_embeds. It consisted of a stack of transformers, a dropout layer, followed by a linear classification head projecting to the number of combined classes. Training was carried out using AdamW optimizer with an initial learning rate of 10e-6, a linear learning rate scheduler with no warmup, and cross-entropy loss for 15 epochs. A stratified 80:20 train-test split was used to retain class balance.

Model performance was assessed by tracking training and validation loss and accuracy at every epoch, with results plotted on convergence graphs. Prediction power at the very end was then assessed through a heatmap of the confusion matrix and an in-depth classification report that encapsulated precision, recall, and F1 scores for every stress-week condition.

Table 4.1 Key hyperparameters and configurations used in this BERT based pipeline.

Component	**Setting / Value**
Transformer backbone	BERT (bert-base-uncased)
Inputs	MFCC + spectral feature embeddings
Output classes	Combined Week_Stress labels
Batch size	16
Epochs	15
Optimizer	AdamW
Learning rate	1e-5
Scheduler	Linear scheduler (no warmup)
Loss function	CrossEntropyLoss
Dropout	BERT default (0.1)
Train:test split	80:20 stratified
Device	CUDA-enabled GPU (Google Colab)

Chapter 5 – Results and Discussions

5.1 Classical ML approach

The statistical approach incorporated an array of supervised learning models; each selected for its potential to deal with the high-dimensional and noisy nature of the acoustic data. Tree-ensemble methods such as Random Forest, Extra Trees, and Gradient Boosting all perform the training phase by constructing various decision trees and combining their outputs during testing to minimize overfitting and maximize robustness; the combination procedure involves either averaging for regression or voting for classification. AdaBoost builds weak learners sequentially in such a way that greater emphasis is placed on samples misclassified at previous iterations, thus allowing higher accuracy to be achieved in subsequent iterations. HistGradientBoosting develops histograms on feature values to speed up training and reduce memory consumption, while maintaining the ability of good performance on big tabular datasets. CatBoost is one gradient boosting method that is well suited for categorical features, as it uses ordered boosting, built-in regularization to avoid overfitting, and also provides model interpretability by outputting feature importance scores.

Neural architectures were investigated with ensembles too. MLPClassifier (Multi-Layer Perceptron) is a feedforward artificial neural network capable of adapting through layers of interconnected neurons and non-linear activation functions to represent any complex nonlinear relationship between features. TabNet, a deep learning model designed for tabular data, applies sequential attention to decide which features it should focus on at each step in its decisions. Its sparse attention masks make it interpretable without sacrificing power of representation and construction of deep learning. For longitudinal modeling, an LSTM network was used to analyze temporal changes in stress vocalizations over weeks. LSTM, a kind of recurrent neural network, retains memory over long sequences and is hence suitable for picking out nuances of vocalization and habituation trends in time-series data.

5.1.1 Health Classification Based on Vocalizations

The complexity of this issue is demonstrated using different models in the study (Table 5.1). The models CatBoost and HistGradientBoosting, as shown by the results in the confusion matrices and all the accuracy plots, are more than capable for the task as they were able to perform well. Both excel at non-linear and feature interaction comprehension, which is crucial in biological datasets. These models have high accuracy and recall which suggests that the data patterns that distinguish healthy and unhealthy birds are captured by these models.

Table 5.1: Performance metrics (accuracy, precision, recall, F1-score) of baseline models for Dataset -1

Model	Accuracy	Precision	Recall	F1 Score
Random Forest	0.977	0.978	0.977	0.977
Gradient Boosting	0.973	0.974	0.973	0.973
AdaBoost	0.988	0.989	0.988	0.988
Extra Trees	0.988	0.989	0.988	0.988
HistGradientBoosting	0.992	0.993	0.992	0.992
CatBoost	0.985	0.985	0.985	0.985
MLPClassifier	0.988	0.989	0.988	0.988
TabNet	0.612	0.616	0.612	0.595

The fact that HistGradientBoosting provides perfect classification accuracy as also per the confusion matrix proves its better performance for both categorical and nu-merical feature sets as well as its ability to adjust the tree-based learning structures for better performance. Compared with others, TabNet did not seem to perform well, and this was because it employs a deep learning model, as such models tend to work best with larger datasets. Because of the size of the dataset, TabNet was not able to generalize well, even though it was very good at learning features. Alternatively, this does not discount its usefulness when applied to more significant datasets or cases with varying feature representations.

From the feature importance analysis in figure 5.1, it is easy to determine what sound subsets of MFCCs were important and feature 7 received the highest importance of 8.8%. This is consistent with the biological knowledge that states how stress or illness changes certain aspects of vocalization. We know MFCC features are representations of the spectral content of vocalization, and their change usually results from a variance in a bird's respiratory and vocal apparatus.

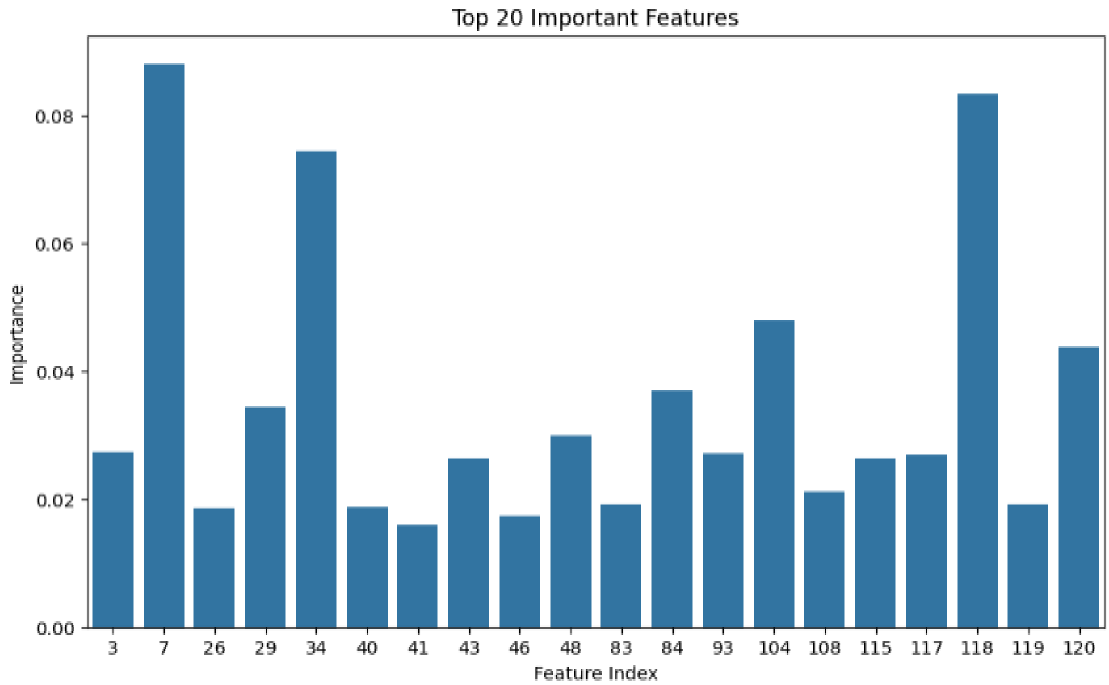

Figure 5.1. The top 20 most important features (left) ranked by their contribution to classification performance, with MFCC-based features showing the highest influence.

For instance, 40 MFCC coefficients were computed per frame with the Librosa library in the dataset we worked with. For each audio clip, MFCC_0 values were averaged across frames to yield a single value per clip. Later, these obtained clip-level MFCC_0 values were pooled by health status, and the mean computed for each group. The results showed that, sick birds registered an average MFCC_0 of −319.2, while healthy birds had an average of −337.6. A two-sample t-test showed that the difference was statistically significant ($p < 0.05$), indicating that the vocal impairment owing to sickness can affect the spectral envelope of the sounds produced. Our results are consistent with previous studies indicating that MFCCs are non-invasive measures of stress or disease in animal vocalizations [32].

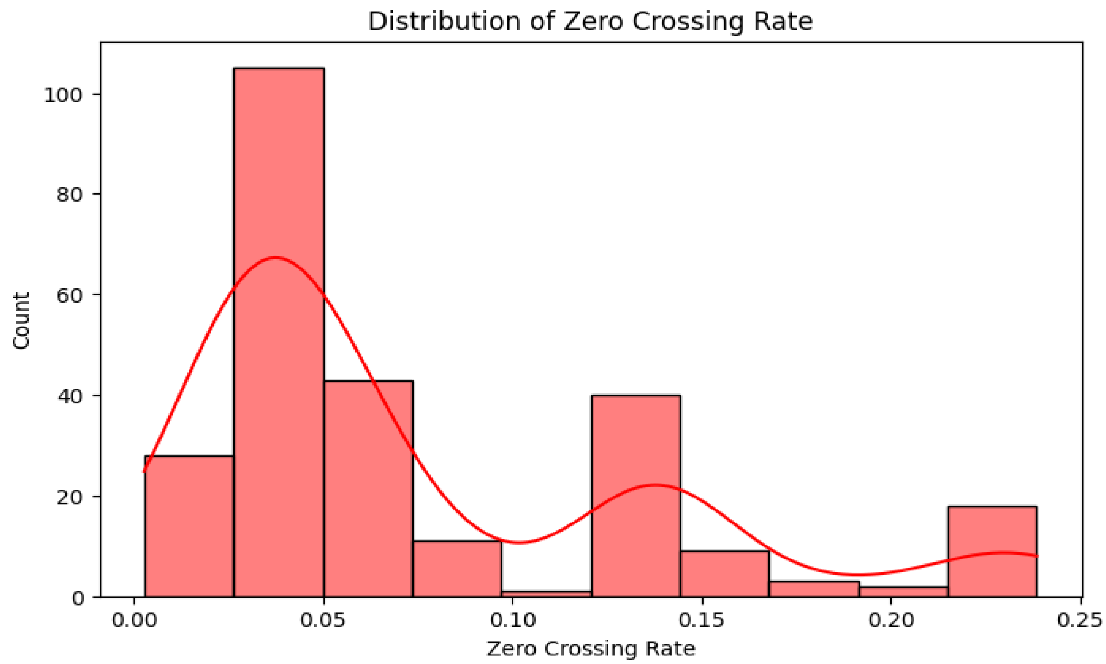

Figure 5.2. Distribution of Zero- Crossing Rate (right) across samples reveals increased variability and skewness in vocalizations.

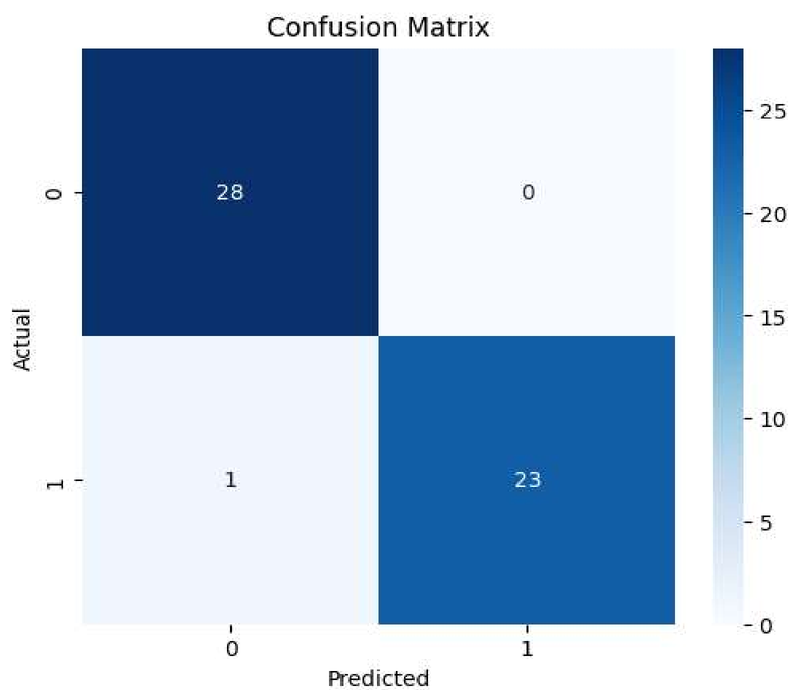

Figure 5.3. Confusion matrix for health classification using the Random Forest machine learning model.

The Zero-Crossing Rate (ZCR) showed more dispersion among unhealthy birds indicating irregularity in the signals or instability in vocalization as seen in Figure 5.2. Using Gini importance averaged over five-fold cross-validation measures, these features were ranked and also confirmed through statistical comparisons that emphasize variation across health conditions. Most of the top 20 consisted of low-order MFCC mean and standard deviation measures (e.g., MFCC_2-MFCC_8), along with spectral contrast means and standard deviations (from bands 0 to 6), and signal-level descriptors, such as Zero-Crossing Rate (ZCR). The confusion matrix for the random forest model is seen in Figure 5.3.

These cover mel-scale energy distribution, harmonic structure, and waveform irregularity, respectively, and ranked high across the folds. A different, but equally important feature, spectral contrast, was significantly lower in unhealthy birds (15.12 vs. 15.95, $p < 0.05$). This parameter measures the degree of sharpness and clarity of a sound's frequency spectrum. The reduction in spectral contrast in unhealthy birds indicates some deterioration in the harmonic structure of their bird songs.

Biologically, this may occur due to respiratory problems or other pathological conditions affecting the larynx, resulting in a greater degree of blur of distinctly recognizable harmonic tones. This can be accounted for by strained breathing or turbulent breathing patterns due to respiratory illness or any other medical condition. The responsiveness of ZCR captures the overall variability in sound production, which serves as a tell-tale in estimating the physiological demands and stress levels sustained by the birds.

5.1.2 Behavioral Vocalization Classification

Chicken vocalizations obtained from ChickenLanguageDataset were classified using five- folds cross-validation as well as by comparing several metrics for measuring and comparing different models for assessing the generalization and robustness of the models. The Behavioral call types studied were "Do_you_have_food", "eating", "greeting", "hungry", "tidbitting_hen", and "where_is_everyone", and are associated with different physiological or social contexts. As seen in Table 5.2, overall classification accuracy for the model that performed best using HistGradientBoosting attained 80.4% and a robust macro F1-score of 0.807, which indicates an appropriate balance in sensitivity and precision across all the types of vocalizations.

Confusion matrices of the top models (Figure 5.4), Extra Trees and HistGradientBoosting, demonstrated a strong classification capability of "tidbitting_hen", "where_is_everyone", and "hungry" categories with little or no misclassifications. "Greeting" and "eating" are rated for moderate confusion with respect to adjacent behavioral labels in the context, showing that the two calls share a considerable amount of acoustic characteristics. Thus, Extra Trees had total classification in "Do_you_have_food" and "hungry" but could not perform well in "greeting," whereas HistGradientBoosting showed stronger generalization across all categories.

Table 5.2 Performance metrics (accuracy, precision, recall, F1-score) of models for Dataset 2

Model	Accuracy	Precision	Recall	F1 Score
Random Forest	0.804	0.803	0.804	0.802
Gradient Boosting	0.607	0.591	0.607	0.591
AdaBoost	0.464	0.533	0.464	0.457
Extra Trees	0.768	0.770	0.768	0.766
HistGradientBoosting	0.804	0.817	0.804	0.807
CatBoost	0.196	0.863	0.857	0.858
MLPClassifier	0.741	0.756	0.741	0.745
TabNet	0.214	0.291	0.214	0.193

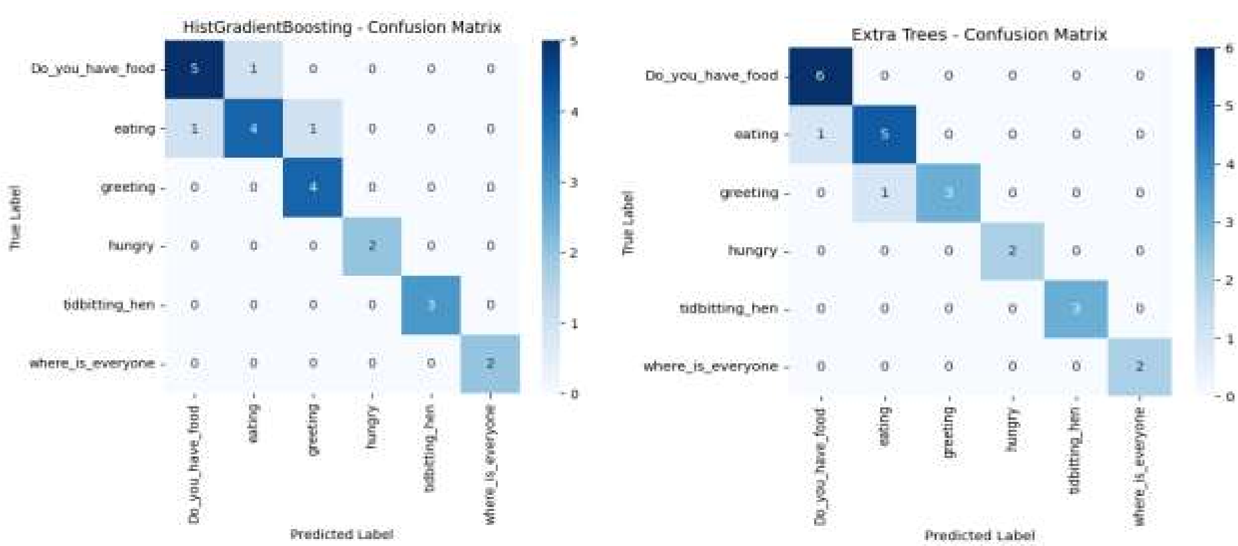

Figure 5.4 The confusion matrices for the top performing Models of the Dataset 2.

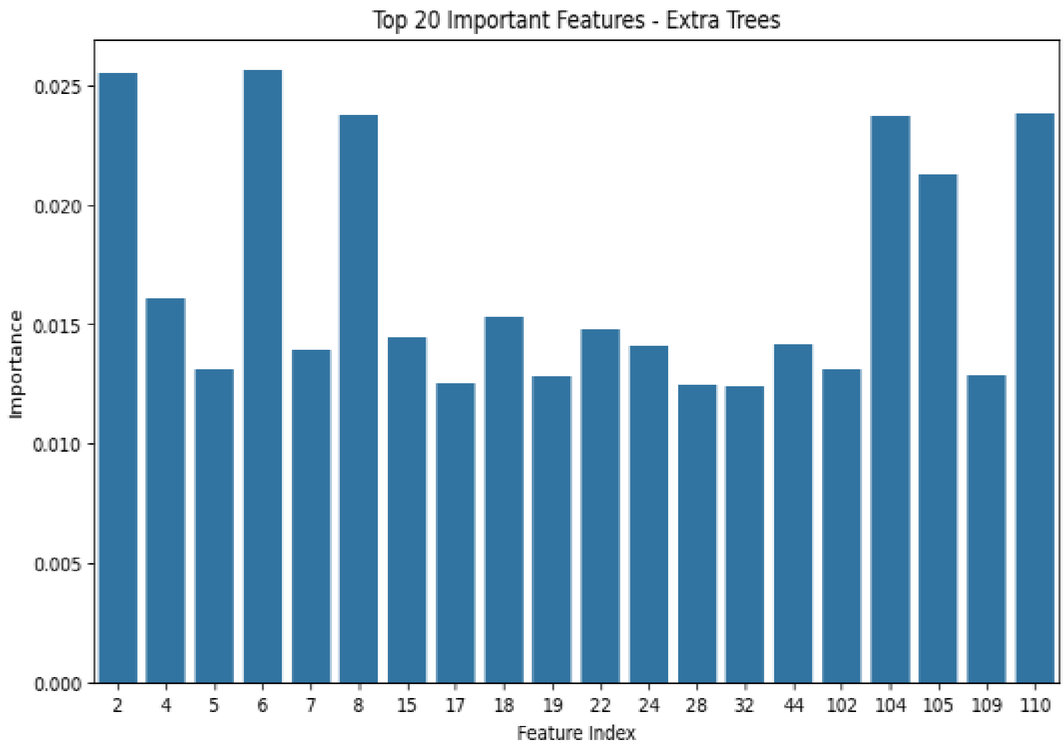

Figure 5.5 Top 20 feature importance plot for Extra Trees

Feature importance analysis (Figure 5.5) revealed that several low-order MFCC coefficients (e.g., Features 2, 4, 6, 7, and 8, corresponding to MFCC_2 to MFCC_8 means) were among the most informative acoustic features. These coefficients usually capture mel-frequency spectral envelope variations between closely linked vocal resonance and energy patterns in animal vocalizations. This included the MFCCs, with the Spectral Contrast Bands 0, 1, and 6 (Features 104, 105, and 110) also, listed among the top contributors. In fact, MFCC features showed clear dissimilarity between different vocalizations in our dataset, which is in line with their established role in detecting distress calls in poultry [43]. The dimensional timbral richness and harmonic structures allowed models to differentiate complex social calls from simple feeding cues. Correlation heat maps for MFCCs show moderate to low redundancy in most features, indicating efficacy in the spread of features without serious collinearity.

Several performance criteria beyond simple raw accuracy were used for the statistical evaluation across the models, including Cohen's Kappa (Figure 5.6), Matthews Correlation Coefficient (MCC) (Figure 5.7), and Log Loss Difference (Figure 5.8). HistGradientBoosting and CatBoost scored the most on agreement (Cohen's Kappa > 0.83).

This shows the reliability of the models against class imbalance. Furthermore, the very high MCC values of above 0.89 for these models confirmed their predictions would be very reliable, particularly in categories of low sample size. MLPClassifier initially showed promise but performed poorly after cross-validation (Cohen's Kappa = 0.18, Log Loss Difference = −6.11), suggesting possible overfitting to training distributions during cross-validation or miscalibration on the minority classes. Overall, log- loss differences between the models emphasized the robustness of HistGradientBoosting and Extra Trees against calibration.

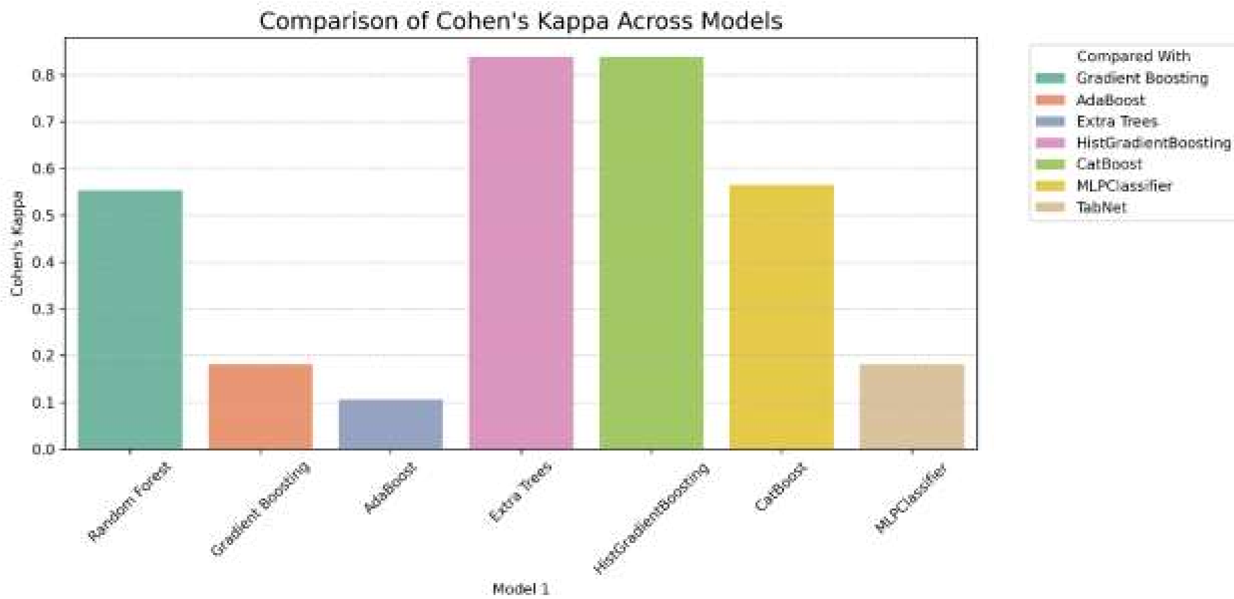

Figure 5.6 Statistical comparison of classifier performance - Cohen's Kappa

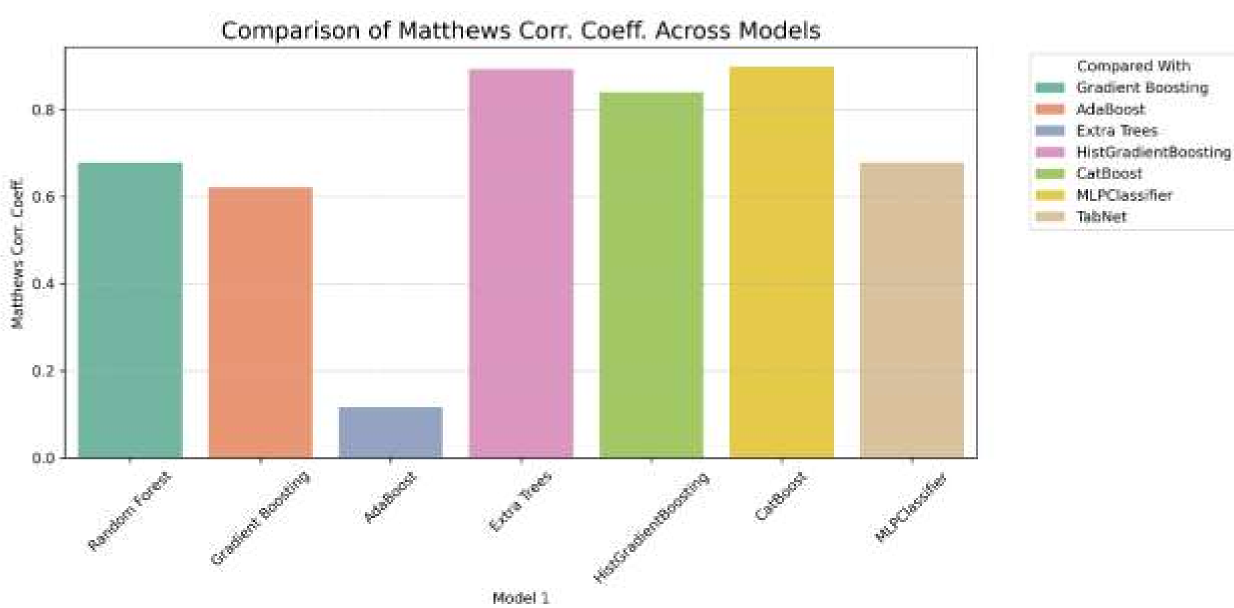

Figure 5.7 Statistical comparison of classifier performance - Matthews Correlation Coefficient

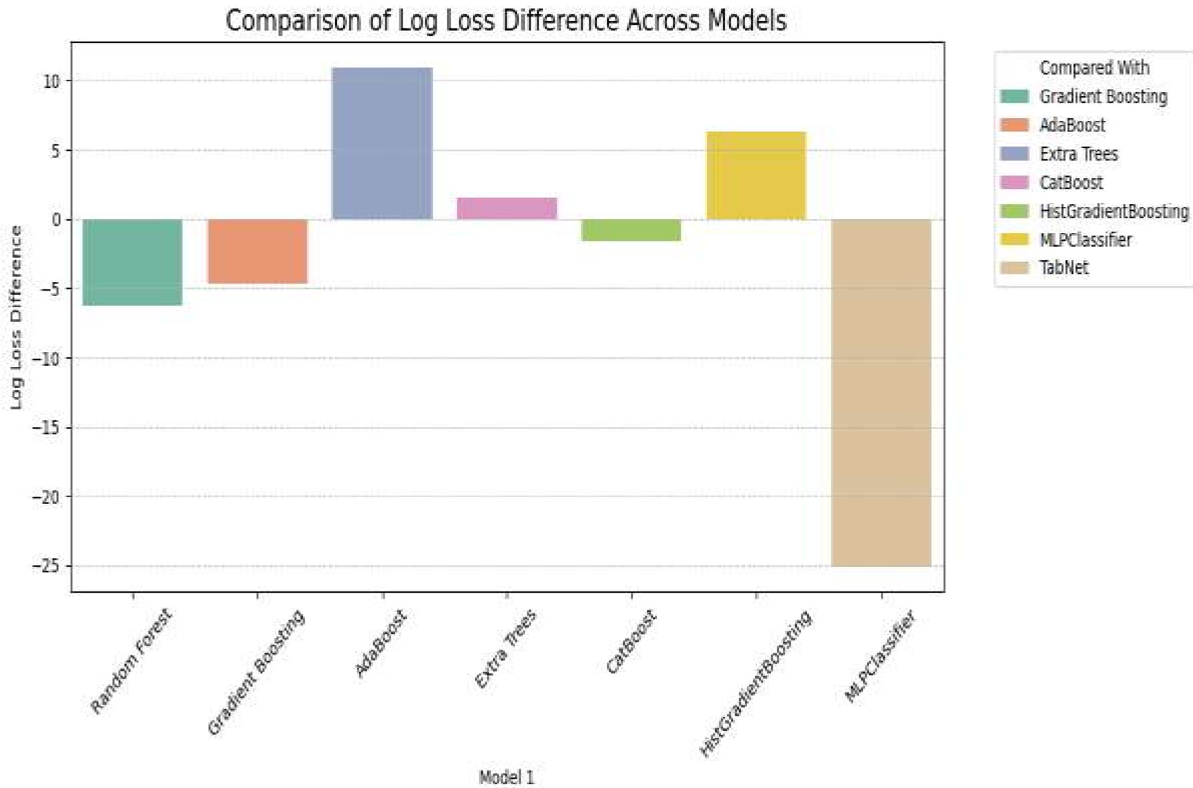

Figure 5.8 Statistical comparison of classifier performance - Log Loss Difference

Cohen's Kappa, MCC, and Log Loss were the statistical comparisons among models represented pictorially in Figure 5. The agreement among models and across statistical measures was further evaluated in a correlation heatmap. The moderate to strong positive correlation between MCC and Cohen's Kappa ($r = 0.75$) suggested similar directions regarding reliability-based metrics. On the contrary, a negative correlation between MCC and Log Loss ($r = -0.24$) indicates that there is a trade-off between the penalties made for a misclassification instance and the confidence of that decision. Such findings thus provide multi-pronged interpretations of classifier robustness.

These MFCC feature correlation plots further proved that key features across samples maintained moderate independence, especially indices 2, 6, 8, 102, and 104, in accordance with biological relevance and the feature selection by tree-based models. Given the varying social and distress calls of chickens, the spectral and cepstral characteristics were important for effectively classifying between vocal behaviors. These results endorse the utility of high-end ensemble classifiers and MFCC-derived features for non-invasive, scalable behavioral monitoring in poultry.

5.1.3 Temporal Trends and Stressor-Specific Vocalization Dynamics

While static classification of health and behavior was done, a longitudinal study was also conducted to observe whether the vocalization pattern changes over time when faced with different stressors. This section is devoted to Dataset 3, wherein week-wise vocal recordings were collected for chickens exposed to visual (Trt1: umbrella opening) and auditory (Trt2: dog barking) stress. Unlike discrete labeling tasks, this setting traces the acoustic dynamics of stress and observes how feature changes over different times, making the features shift or stabilize or persist as viable acoustic features.

The shifts are modeled with LSTM networks and contrasted with feature trends to discover habituation, individual variation, and stressor-specific adaptations embedded in the vocal signals. These results help situate how chickens respond physiologically and behaviorally to repeated environmental challenges, which sets the premise for real-time welfare monitoring under dynamic farm environments.

One of the most highlighted observations was the separation trends between the two treatment groups: Trt1 and Trt2 as seen in figure 5.9. This spreading indicates the inherent differences in the perception and effect of chicken species against different types of stressors. For example, Trt1 (opening of an umbrella) is a type of stressor that is visually induced. The response is normally characterized by surprise and triggers an acute response. Often these responses are defined by quick, sharp changes in vocalization patterns, which is indicated in the initial weeks of the study.

Contrarily, Trt2 (dog barking) is an auditory stressor that is more sustained, and in this instance, the vocalizations are reflective of a prolonged adaptation process. Auditory stressors engage different sensory and neurological pathways compared with visual stressors, consequently, yielding distinct responses.

This differentiation illustrates that when analyzing vocalization data, the nature of the stressor has to be accounted for, as individual animals may be differentially responsive based on the type and timing of stressors. Variations in the traits of vocalizations over weeks also portray the shifting nature of stress responses. The sharp oscillations in many of these features during these early weeks suggest that the response is possibly driven by the novelty of the stressors and hence, the sensitivity of the birds during these first encounters.

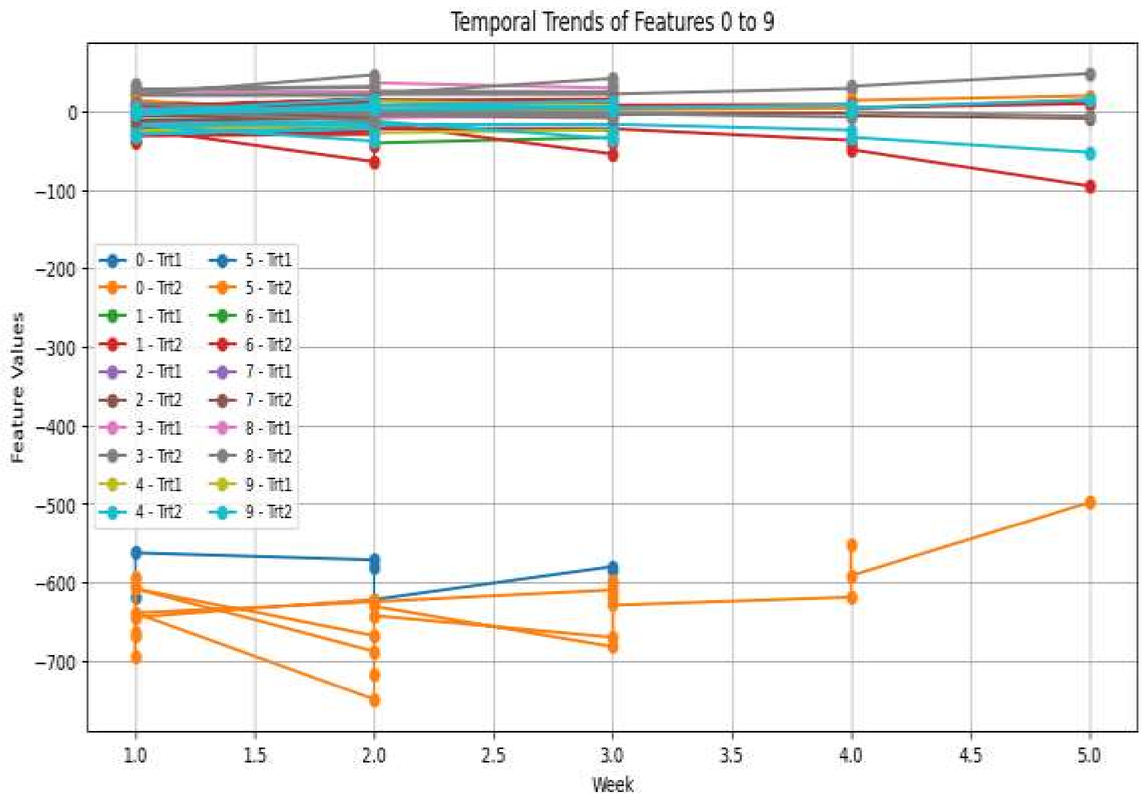

Figure 5.9. Temporal trends in acoustic features across weeks for Trt1 (umbrella opening) and Trt2 (dog barking). The plot (features 0–9) shows early oscillations and subsequent stabilization

After some time, the trends stabilize away, suggesting a process of habituation. Habituation is a well-studied phenomenon in animal behavior: Each time an animal confronts the same stressor, its response diminishes. This has also been interpreted to indicate reduced stress levels on the part of the chickens since they have become accustomed. Alternatively, it may be interpreted as physiological responses adjusted to cope better with stress. Such adaptations are crucial for survival, more so in commercial poultry farming, where animals are often faced with different stress-inducing stimuli. That being said, while some features have settled down with time, others show a continued high standard of variability. Constant variability may indicate individual variability within birds, as not all animals respond to stressors in the same manner. Some of the most important factors that work to influence an individual's response to an environmental challenge include genetic endowment, age, social rank, and past experiences. Such differences must be taken into account when devising any stress-reducing interventions, as any given intervention may not be equally effective for the entire herd of animals.

The further grouping of features into segments for analyses offers further revelations concerning the duplicity of vocalization patterns. Thus, averaging the values within the feature groups reveals much broader trends that otherwise would have been shrouded in noise from the individual features. The Segregation of Trt1 and Trt2 through these groups further strengthens that vocalization patterns reflect stress-or-specific adaptations. In addition, because of grouping, the analysis of high-dimensional data is spanned out better; thus, researchers may focus on the most relevant patterns.

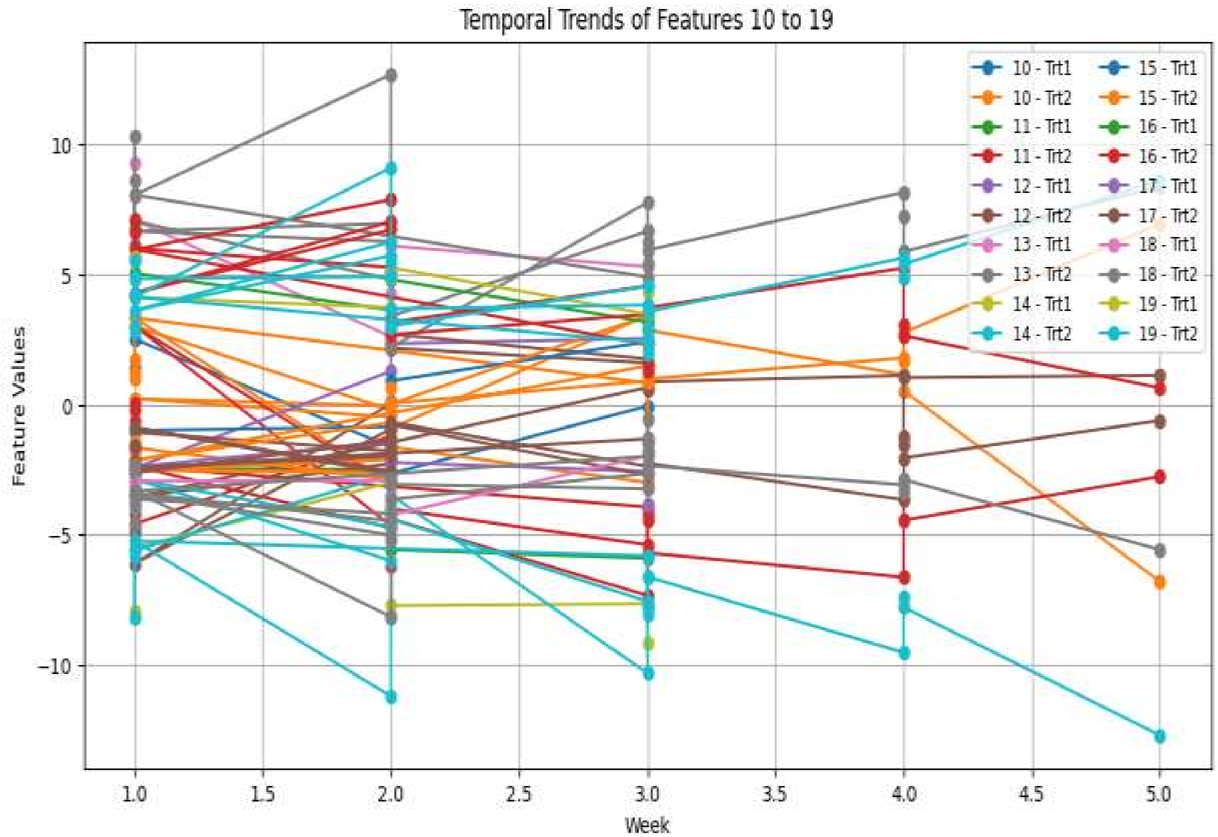

Figure 5.10 The plot (features 10–19) reveals persistent variability, indicating stressor-specific and individual response patterns.

These trends may reflect underlying physiological mechanisms of vocal control, particularly autonomic modulation of the syrinx. Vocalization is mediated through the syrinx, the avian voice organ, which is under autonomic control. During stressful periods, there are changes in the autonomic balance with corresponding changes in pitch, frequency, and amplitude. More than that, these changes are not random, because they reveal an animal's internal state. For example, being higher-pitched or more frequent evokes a picture of heightened arousal or misery. The most distinct trends in Trt1 and Trt2 are likely to disturb the congruency of how these stressors affect the autonomic regulation of the syrinx.

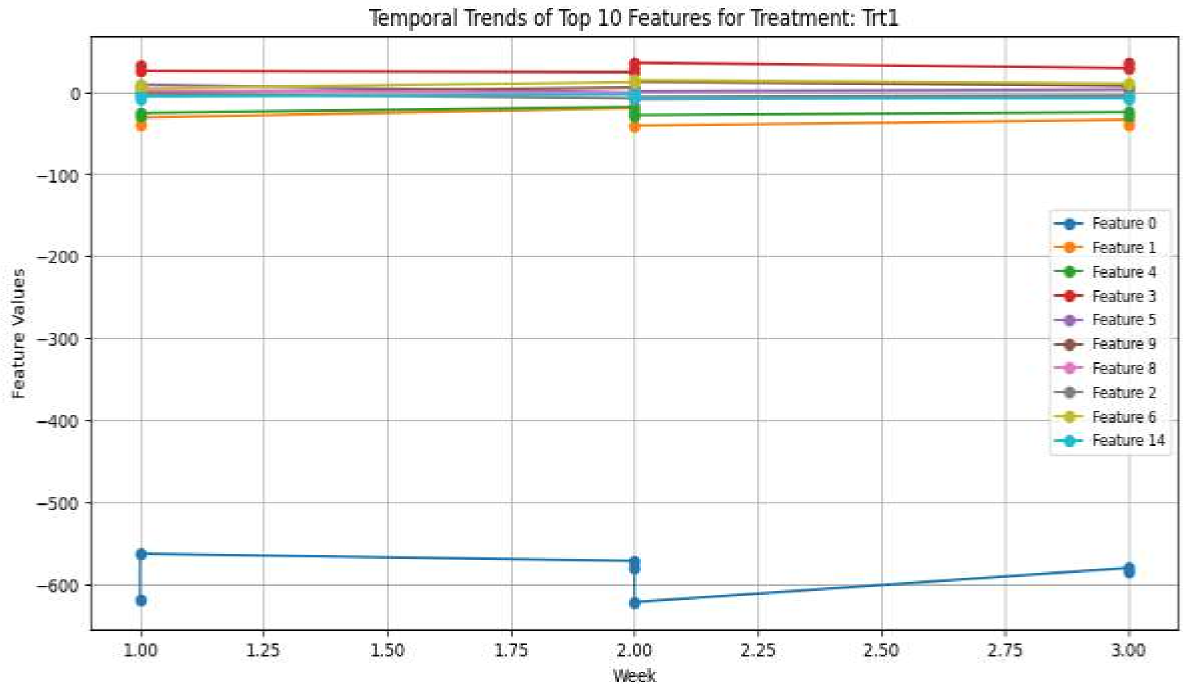

Figure 5.11 Temporal trends in selected acoustic features under individual treatments (TRT 1)

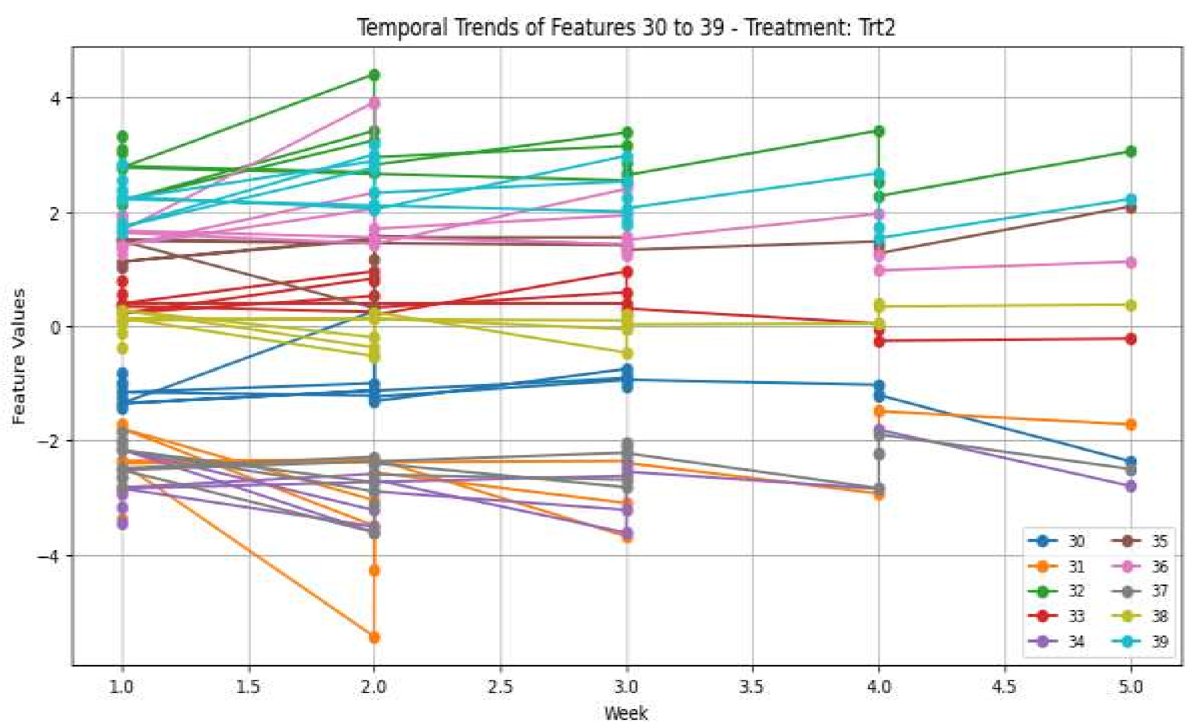

Figure 5.12 Trt2 elicits higher variability and prolonged divergence across features

Adding to this is the time- embedding in a dataset: by introducing times, more complications arise. Many factors, including the age and developing stage of the birds, might affect the week-to-week changes in vocalization features. In younger birds, vocalization is much more variable owing to their relatively undeveloped or immature stress handling and behavioral acclimatization. This is probably why the more volatile feature trends were observed during the earlier weeks of these plots.

The more the young birds grow, the more control they have over their reactions, perhaps because their systems develop cognitive profiles to cope with stress well. The flock social dynamics that might enable that can be a variable, i.e., chickens are social, and their vocalizations function along that path of communication.

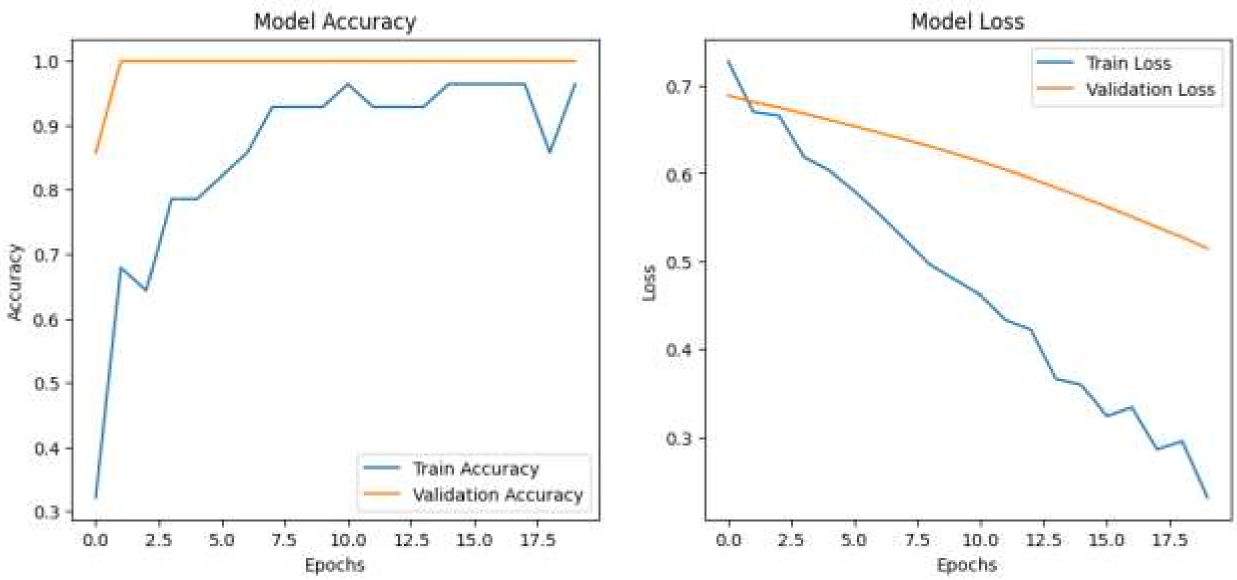

Figure 5.13 Training and validation performance of the LSTM model for stress classification in poultry.

We employed an LSTM-based deep learning model to assess stress-induced vocalization patterns under two experimental conditions: visual stress (e.g., umbrella opening) and auditory stress (e.g., dog barking). Extracted MFCC features from preprocessed audio clips were normalized and reshaped for time series modeling. The architecture consisted of two LSTM layers (128 and 64 units) with batch normalization and dropout layers for regularization, followed by dense layers using ReLU and sigmoid activations for binary classification. The training was carried out utilizing the Adam optimizer with binary cross-entropy loss, applying early stopping and learning rate scheduling strategies based on validation loss.

For 20 epochs, the model converged rapidly, achieving high accuracies (greater than 95 percent) for both training and validation curves (Figure 5.13), and showed no signs of overfitting as validated by steadily decreasing loss values. Based on the biological implications, the performance of the model indicates clear acoustic variation between vocalizations depending on the types of stress applied. The high accuracy further suggests that chickens possess structured and distinguishable vocal responses that reflect the context of the stressor. Visual stress would result in an abrupt change in acoustic features typical of acute arousal, whereas auditory stress would result in prolonged variability, likely representative of a longer time horizon for adaptation. These patterns are likely under the control of the autonomic nervous system, with modulation reflected by their emotional or physiological state.

An altered vocalization pattern due to stress can potentially modify flock behavior; an example would be a dominant bird that would voice certain calls to maintain its status, while subordinate chickens might show more stressed vocalizations. Among other reasons, the welfare of the whole flock can indeed change from the chain of cause and effect manifesting from these dynamics, which action points to-ward a more holistic understanding of poultry stress and management. Such different temporal patterns have practical implications for poultry farming.

Monitoring these vocalization patterns provides a non-invasive way to assess the welfare of chickens. Knowing the exact conditions that create vocal patterns of stress will allow farmers to ameliorate them. When, for instance, certain stressors consistently lead to an increase in vocalization variability, it this can indicate that the birds are very distressed. Addressing such a stressor, whether through its removal, environmental change, or manipulation of the way chickens were handled by personnel, is vital for animal welfare and productivity.

5.2 Semantic Approach

Using NLP and transformer models to analyze poultry vocalizations has uncovered profound insights about how stress levels and health conditions coexist with vocal behavior. Different patterns of vocal characteristics have been recorded in birds depending upon their physiological and emotional conditions: pitch, frequency, sentiment, and phonetic structure. In particular, by way of pitch and frequency, higher pitch within a narrow band of frequency was commonly found in the stressed group compared to the relaxed group.

The Wave2Vec 2.0 model could efficiently pick up these subtle nuances, with the stressed cluster tones centered around the higher frequency region (480-500 Hz) safeguarded by physiological constraints of stress such as muscle tension and irregular respiration. On the other hand, unconstrained vocalizations appearing from spirits at ease showed dispersed frequency, translating to larger vocal output freedom. This was further contrasted in terms of sentiments of poultry vocalizations by employing the fine-tuned BERT model.

Prestress sentiment is spread rather evenly, with neutral, positive, and negative sentiments, depicting everyday interactions with mild environmental challenges. The poststress sentiment was increasingly negative with concomitant less neutral sentiment, thus suggesting the enduring effect of stress on the emotions displayed by the chirping of birds. This information proves highly relevant to animal welfare because it opens an avenue for sentiment analysis to be operationalized as a real-time welfare assessment tool and stress detector.

Analysis of phonetic composition gave us deeper insight into the matter of vocalizations. There is domination of short and staccato-like sounds from stressed birds-words which are characterized as consonant-heavy and consist of abrupt and sharp sounds. On the other hand, vocalizations by relaxed birds were more vowel-heavy and thus more melodic and fluid in pattern. These differences give an implication that stressed birds produce vocalizations that are much less complex from a tonal standpoint, either as a result of physiological restrictions or as an adaptive behavior to conserve energy. Due to the vocal emphasis on the consonants, harsh vocalizations may be employed to share distress or to warn flock members.

Changing vocalizations have physiological and behavioral importance. Increased perspiration under Stress may cause an increase in pitch and a decrease in the range of vocalizations, which accompany muscle tension in the laryngeal region caused by stress and increased respiratory rate, common physiological responses to stress in birds.

A narrow pitch frequency and less phonetic variation in poststress vocalizations can be an adaptive energy conservation strategy: the less energy costful the vocalization is, the more the animal can save energy for other vital activities. Behaviorally, a stressed bird might alter its vocal pattern to alert other flock members to distress and maintain cohesion and support in adverse environments. Our results show that vocalizations are useful indicators of stress and health conditions in poultry and thus provide a baseline for non-invasive monitoring for poultry welfare.

5.2.1 Pitch Analysis During Stress Phases

The prestress pitch analysis shows that there was a widespread distribution of pitch frequencies, with a large number centered at the 500 Hz mark as seen in figure 5.14. The pitch histogram indicates an extensive vocal frequency range in the reservoir, implying a calm and steady state where the vocal emissions emanate effortlessly. No clear grouping occurred in the frequency range, suggesting that prestress vocalizations remain in the casual and normal category.

The state of comfort was, thus, conducive to the animals. These vocalizations are probably a sign of communication among the chickens, including a variety of vocal expressions for ordinary activities such as feeding, resting, and exploring. In contrast, the poststress pitch analysis (Figure 5.15) signifies a dramatic narrowing of the pitch frequency range based on clustering nearer to 480-500 Hz.

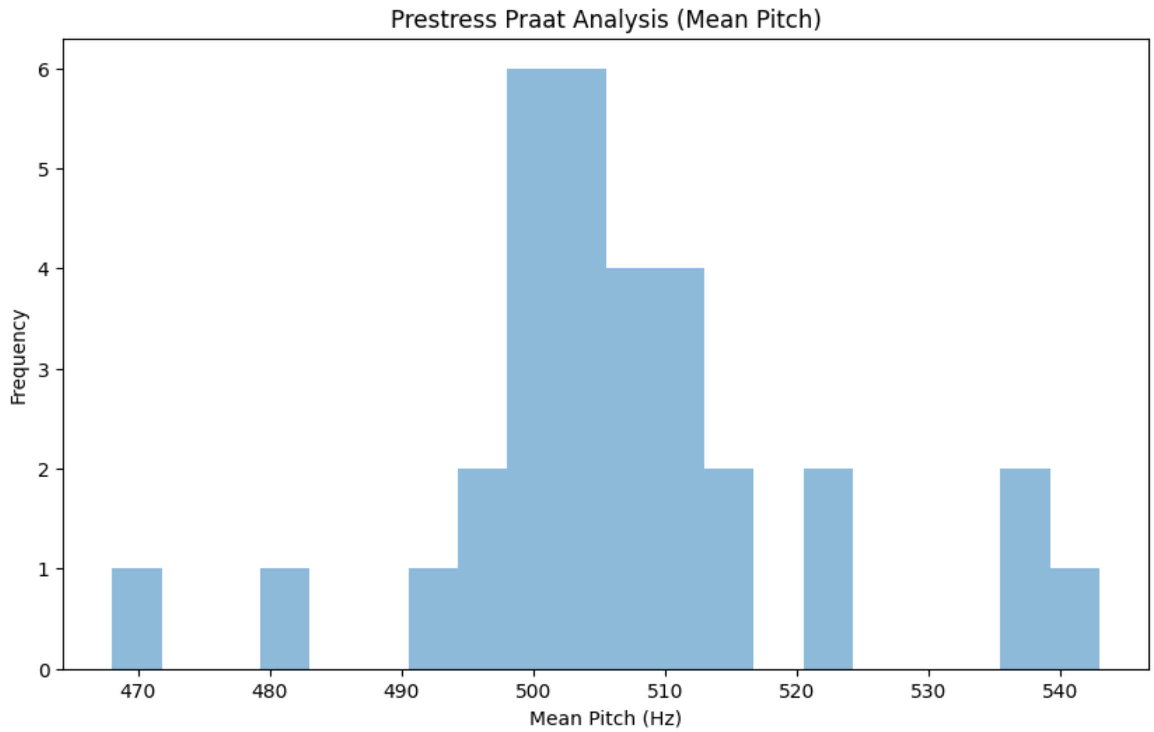

Figure 5.14 Prestress Pitch Analysis. Distribution of pitch frequencies during the prestress phase, highlighting vocal behavior patterns under non-stressful conditions.

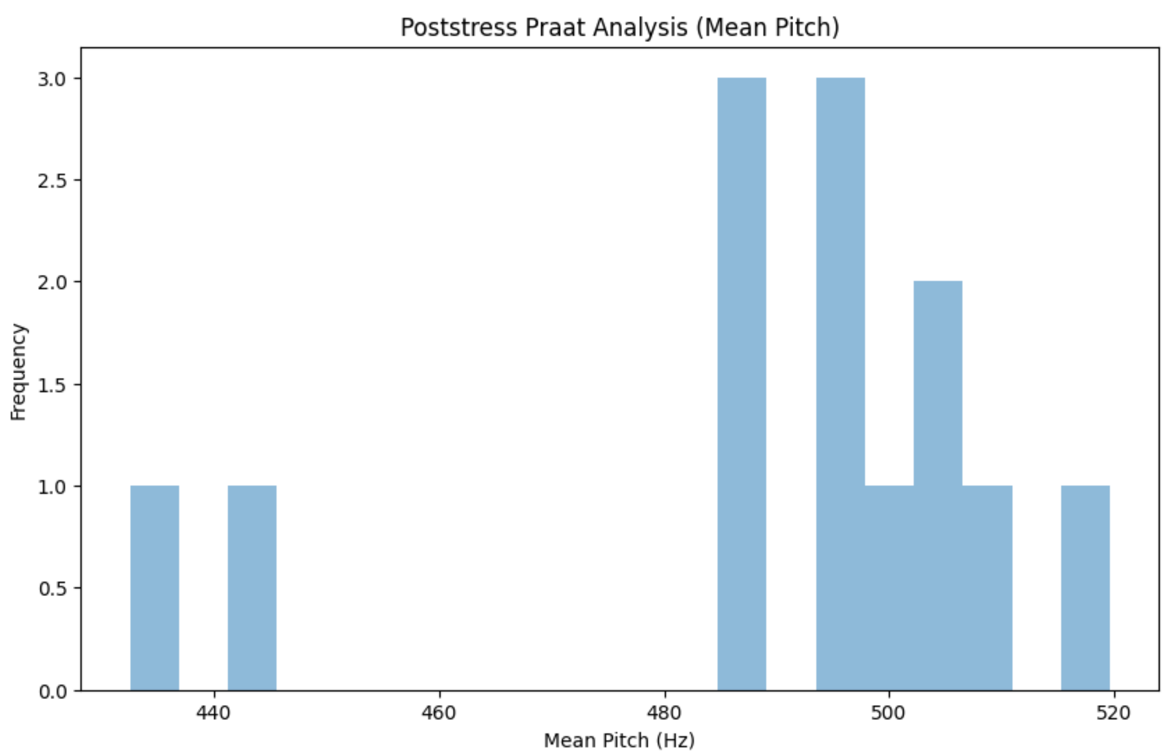

Figure 5.15 Poststress Pitch Analysis. Distribution of pitch frequencies during the poststress phase, indicating changes in vocal behavior following stress.

A decrease in the variety of pitch is suggestive of a physiological restriction to vocalization, possibly due to muscle constraints imposed upon the vocal apparatus by stress. These vocal behaviors might be due to less air flowing through tight throat muscles and chickens trying to save energy. This fits with the idea of adaptive vocal suppression where stressed animals make fewer sounds to save resources or avoid attracting predators.

5.2.2 Physiological and Behavioral Implications of Stress-Induced Vocal Changes

The shifts in pitch frequencies between prestress and poststress phases have essential physiological and behavioral implications. Physiologically, stress translates into muscle tension, especially in the laryngeal muscles involved in pitch control. When it is under stress, these muscles tend to be tightly bound, making the vocal cords unable to vary their pitch effectively, thus narrowing their frequency spectrum. Also, stressed respiration becomes shallow and irregular, which interferes with the airflow required for pitch variation.

This constraint in turn reduces the richness and complexity of vocalizations, giving rise to the monotone vocal profile of stress. Behaviorally, such vocal interference might be adaptive in nature. On encountering stress, animals alter their communication in order to warn others in the group. The tendency to cluster around the mid-range frequencies in the post-stress phase may indicate that chickens attempt consciously to create more distinct sounds that are easily perceived by their flock members.

5.2.3 Sentiment Analysis of the Stress Dataset

We can see the sentiments of prestress phase in figure 5.16. A substantial number of vocalizations lack emotion because neutral sentiments represent 39.3% of the total distribution. Non-stress vocalizations exist in this category, and they resemble sounds which animals produce during social and exploratory encounters. The equal distribution between positive (30%) and negative (30.7%) sentiments suggests a steady emotional balance which probably comes from standard interactions that have small negative elements with positive aspects. Environmental stressors exist in the environment, but their intensity does not reach the point where they create substantial stress according to negative sentiment indicators.

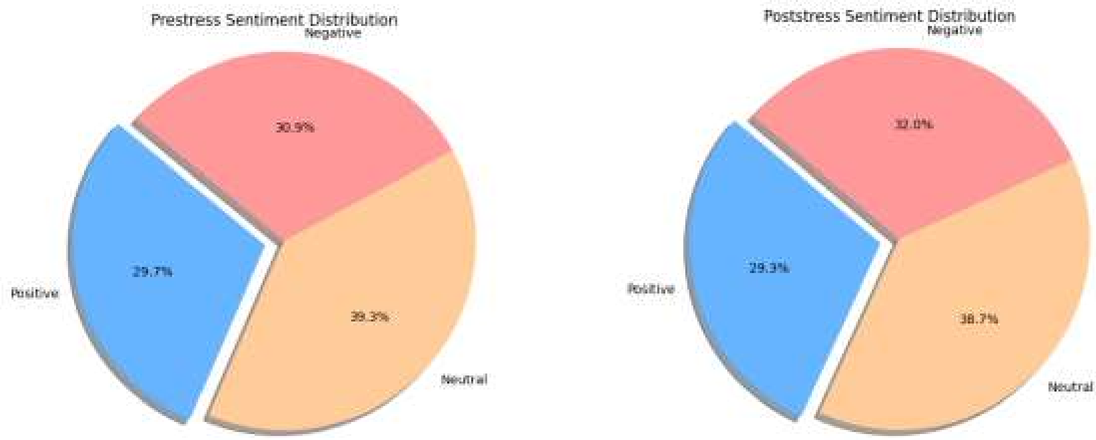

Figure 5.16 Sentiment analysis of poultry vocalizations during the prestress, poststress phase

The emotional response to the post-stress phase has significant changes compared to Prestress phase. The post-stress phase shows a decrease in neutral sentiment to 38.7% alongside a rise in negative sentiment to 32%. The rise in post-stress vocalizations likely represents continued stress-related processes which maintain stress responses while neutral statements refer to recovery attempts from stress-related conditions that continue to show post-stress effects. The positive sentiment score maintains stability at 29.3%. The emotional state of chickens during their post-stress recovery demonstrates emerging resilience through their attempts to return to normal states which differ from their pre-stress emotional condition.

5.2.4 Word Frequency analysis in Stress Dataset

In Figure 5.17, the word frequency distribution is shown to be highly skewed, where relatively few words completely dominate the vocalizations with very many words having far fewer instances. The top-most frequent words (above 500 occurrences) are very simple and common vocalizations believed to be used for some kind of basic communication or signaling. After that, the frequency count drops sharply, indicating that the chickens might have a very large vocabulary, but that most of them are situational or context-specific. The long tail of less frequent words suggests that chicken vocalizations are quite complex and that their vocal repertoire has become adapted to handle many different situations, even if some vocalizations are rarely used.

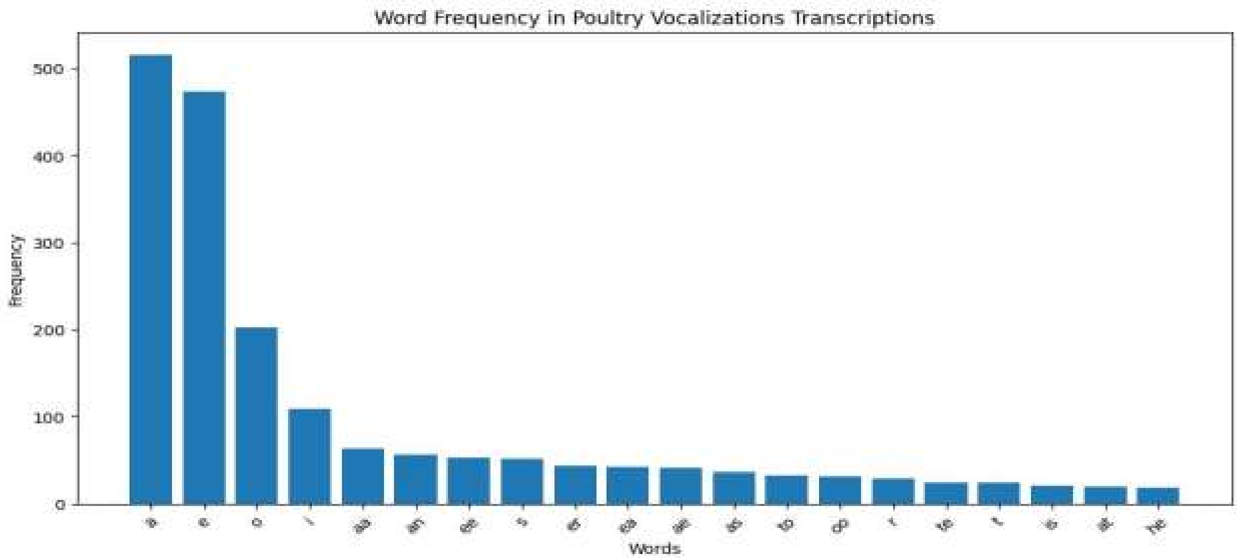

Figure 5.17 Word Frequency Distribution.

The word cloud in Figure 5.18 also shows the prominence of several vowel-rich utterances like "AAA", "EEE", and "OO" through the emotional states. Such common vocalizations might be the nucleus of the poultry communication, which means that they are simple calls that every chicken within the flock equally recognizes. Lesser frequent vocalizations, scattered over the cloud, may denote more specific or detailed communications with a heavier emotional weight or link to an environmental stimulus. Redundancy of loud calls-the most common vocal elements-could give a hint as to their function in maintaining the credibility and effectiveness of communication within the flock during the framework of daily activities or social interactions.

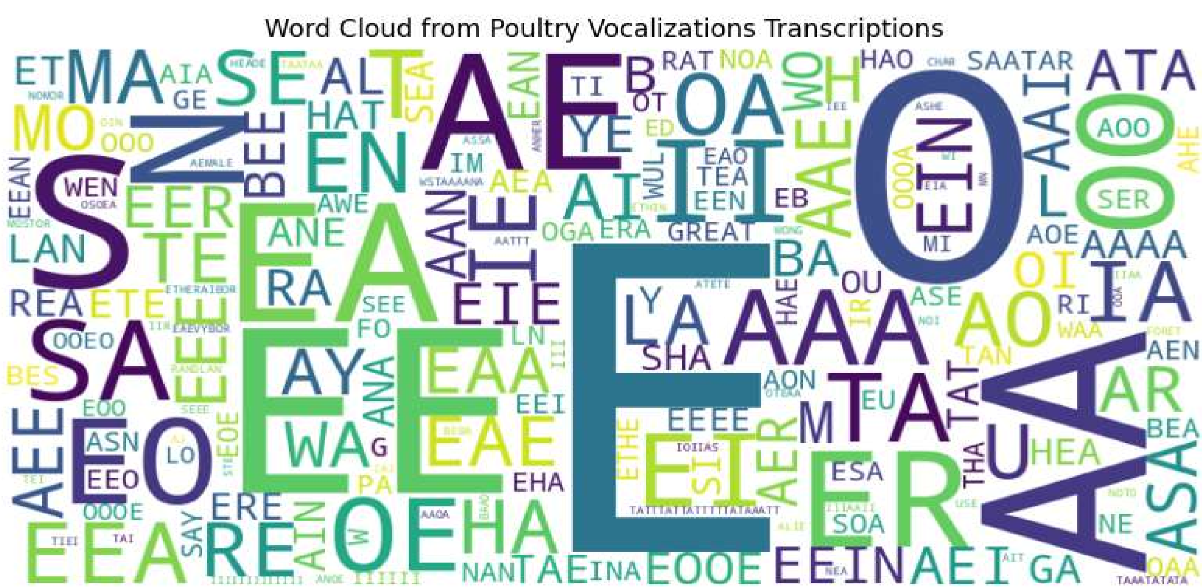

Figure 5.18 Word Cloud of Poultry Vocalizations.

5.2.5 Phonetic Composition Analysis

The phonetic makeup of these vocalizations is shown in Figure 5.19, exhibiting a bimodal distribution. Two clusters are formed: one for vowels and one for consonants. The fact that these clusters exist toward the lower count for both vowels and consonants implies that during the prestress phase, shorter, more concise vocal expressions are often used. The distributions for the vowels and consonants overlap, which means that while most vocalizations are short, there are certain complex ones that lean more toward vowel-rich vocalizations. The presence of the vowels may suggest that there is some element of tone modulation to convey degrees of intensity or intent.

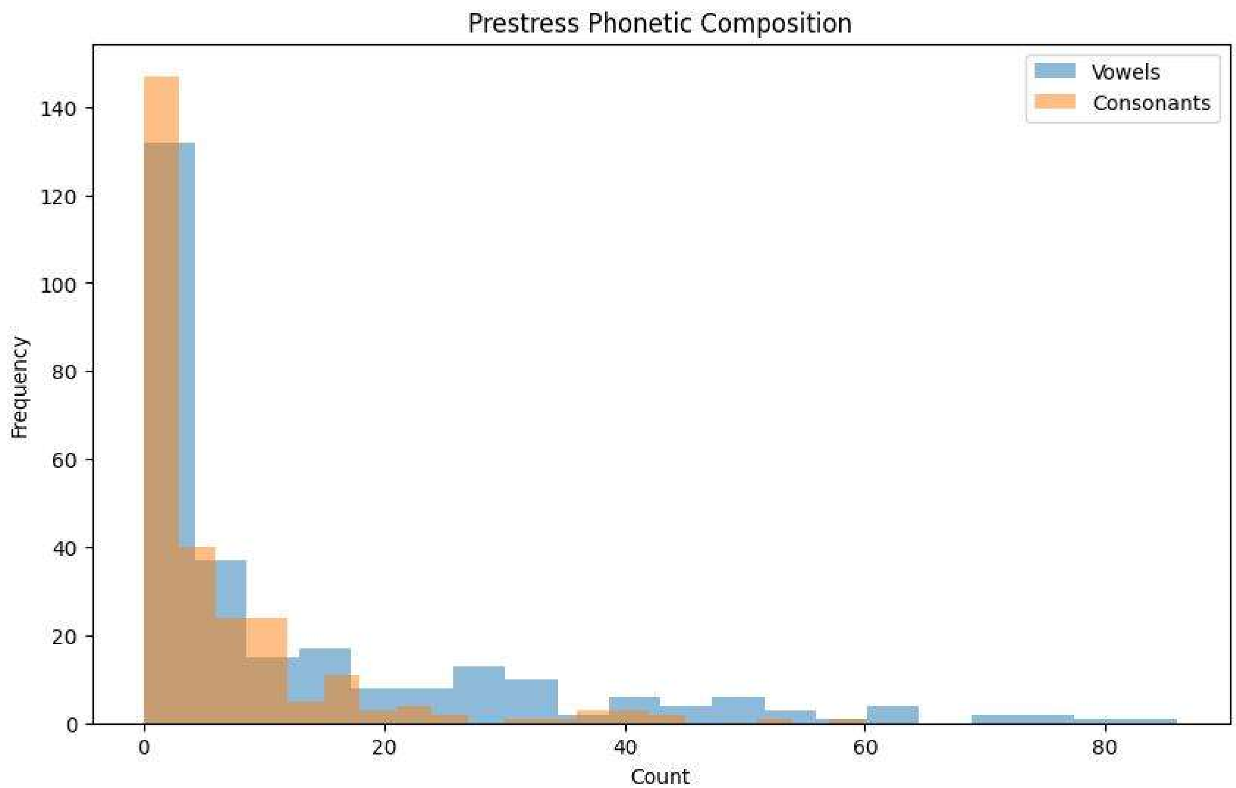

Figure 5.19 Prestress Phonetic Composition.

In the poststress phase, phonetic analysis shows a vowel dominance shift particularly at the higher frequency levels of twenty and above. This suggests that a vocal strategy change has taken place, possibly because of an emotional adjustment following the stressful event. The presence of vowels increases during this phase which could be viewed as an attempt to generate more soothing vocalizations which aids in emotional regulation or contributes towards group cohesion after stress responses have been invoked. The consonant reduction noted during this phase indicates moving away from jarring staccato sounds and supports the idea that poststress utterances are modified to limit vocal sharpness and promote calmness instead.

The phonetic constitution comparison of prestress against poststress status will highlight several predominant tendencies. Given the poststress emphasis on vowels, it points to an emotion-laden transition towards relaxation, whereas lesser anterior consonants could imply lesser meaning or prominence being given to consonants that usually denote harshness with complexity. Given this consonant harshness, their poststress diminution will probably smooth communication from urgent to melodic. The scattered outliers with differing numbers of vowels and consonants in the prestress phase point to a broader vocal repertoire that may link with high levels of emotional arousal or variability in vocal expressions.

5.2.6 Training Dynamics of BERT on Stress Dataset

To classify week-wise stress responses from vocalizations, a supervised classifier was developed using BERT. The model architecture was based on bert-base-uncased from the Hugging Face Transformers library. Instead of processing natural language, the model ingests 768-dimensional acoustic embeddings derived from each vocalization segment. These embeddings, extracted using a speech representation model, were passed directly into the BERT model via the inputs_embeds interface, which bypasses the default token embedding layer. raining was performed using standard cross-entropy loss, learning rate scheduling, and stratified data splits. Evaluation metrics included accuracy, F1-score, and confusion matrices to assess classification performance across stress conditions and temporal stages.

Figure 5.20 depicts the variations in accuracy during training and validation. Both accuracy measures steadily improve across epochs and stabilize just around 93% for validation accuracy. This stability, along with consistent decreases in the values of loss metrics, highlights the ability of the transformer based architecture to robustly learn from complex vocal data. The convergence of the training and validation metrics implies that the model successfully captures the vocal variations caused by stress, thereby asserting its practical implications in real-world domains where stress-induced vocal behavior monitoring is performed.

Meanwhile, Figure 5.21 shows the training and validation losses against 15 epochs for the BERT-based model. The steadied decrease of losses for both training and validation demonstrate that the model is able to learn. Such convergence between two metrics also suggests that the model is well-generalized, without over-fitting, due to dropout regularization and learning rate scheduler. Moreover, the gradual cascades of loss metrics through epochs imply that the model captures meaningful representations of poultry vocalization and subtle emotional dissimilarities.

Figure 5.20 Training vs. Validation Accuracy Over 15 Epochs. Plot of training versus validation accuracy over 15 epochs, showing the model's performance in classifying poultry vocalizations.

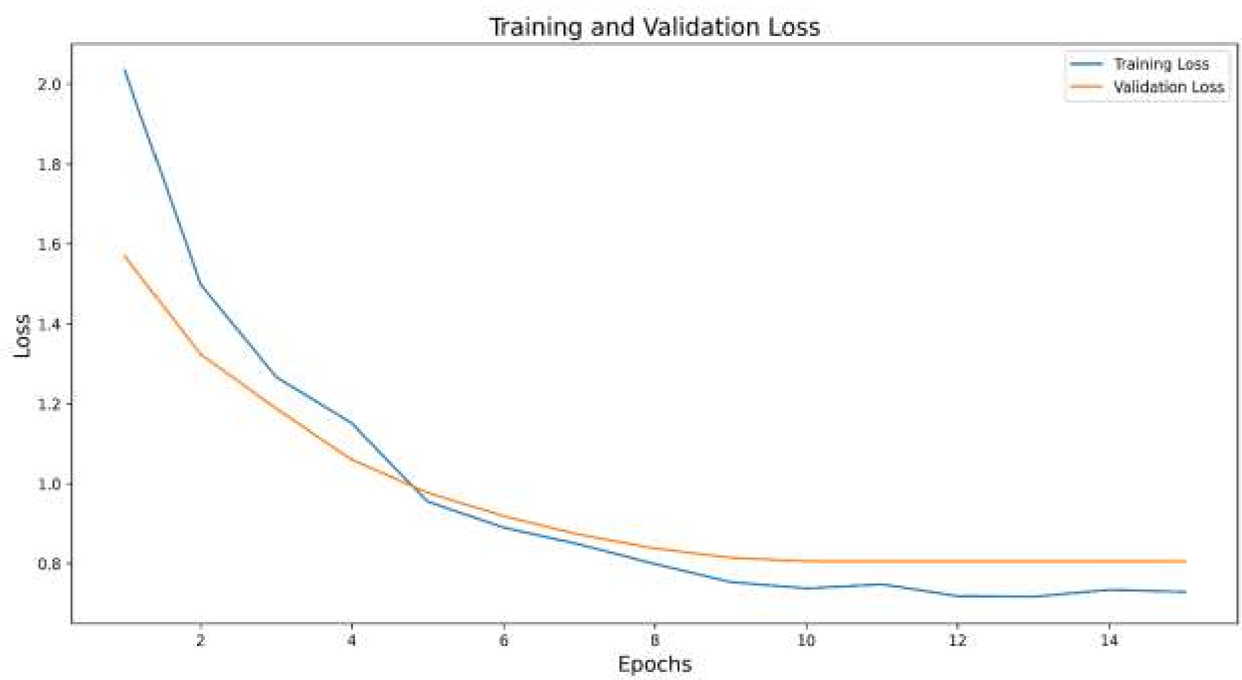

Figure 5.21. Training vs. Validation Loss Over 15 Epochs. Plot of training versus validation loss over 15 epochs, illustrating the model's convergence and generalization during training.

5.2.7 Semantic Analysis of Healthy vs. Unhealthy Conditions

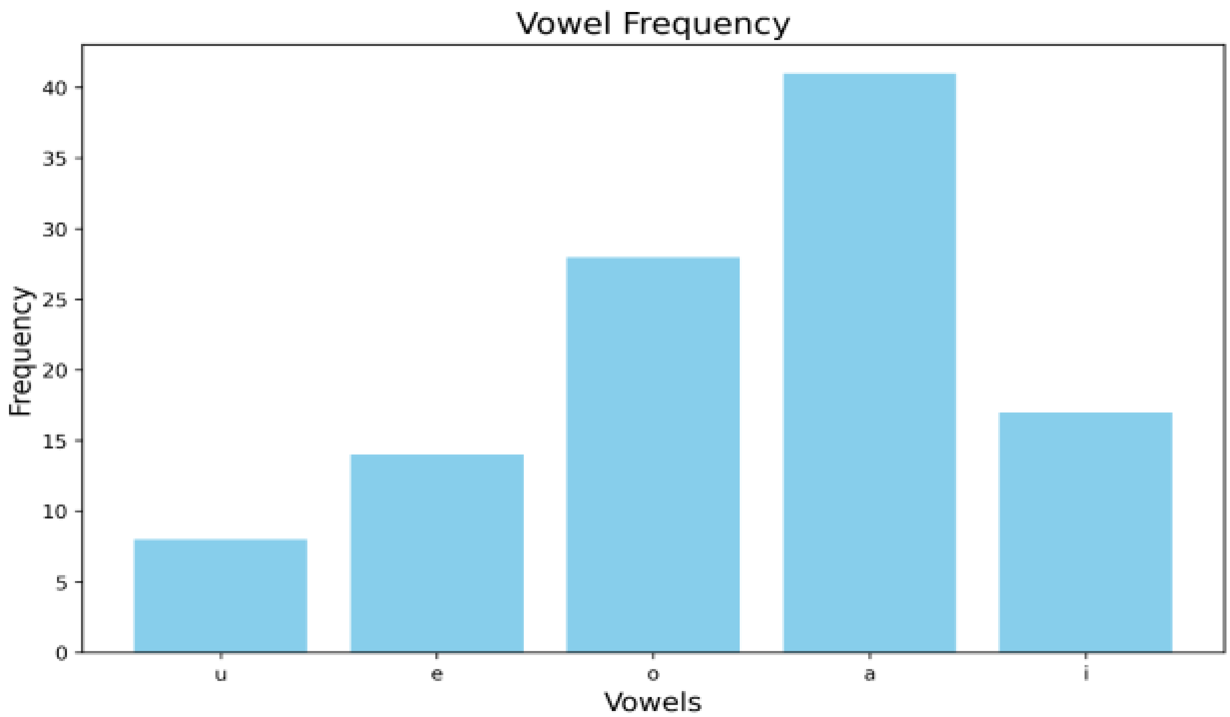

Figure 5.22 Vowel Frequency Comparison showing the differences in vocalization patterns based on health status.

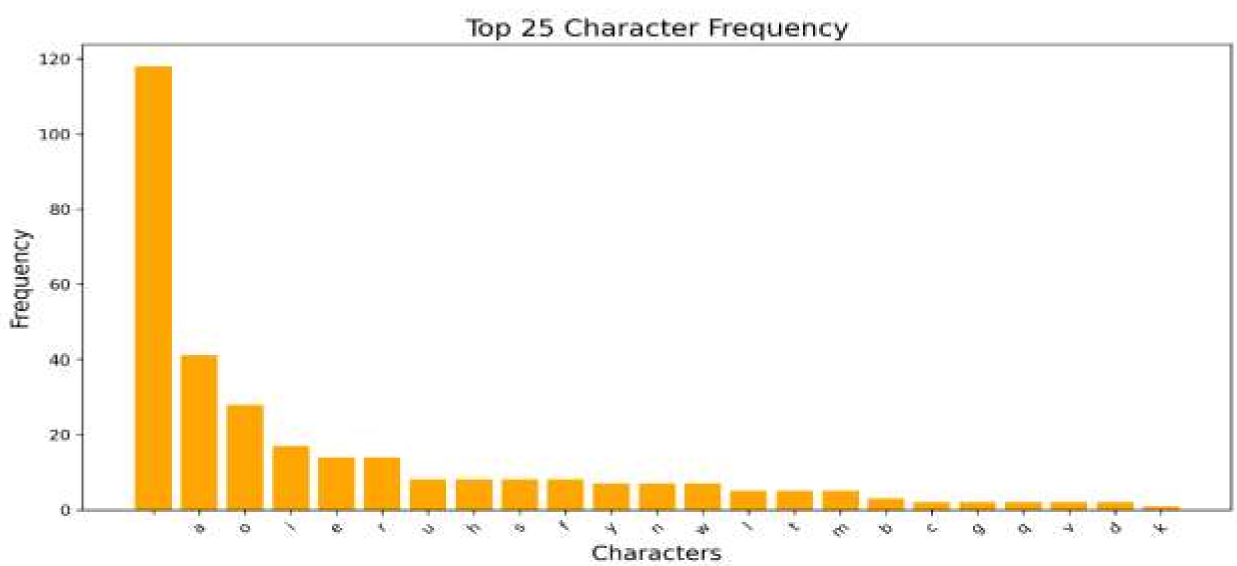

Figure 5.23 Character Frequency Distribution in Poultry Vocalizations illustrating top 25 most frequently occurring characters

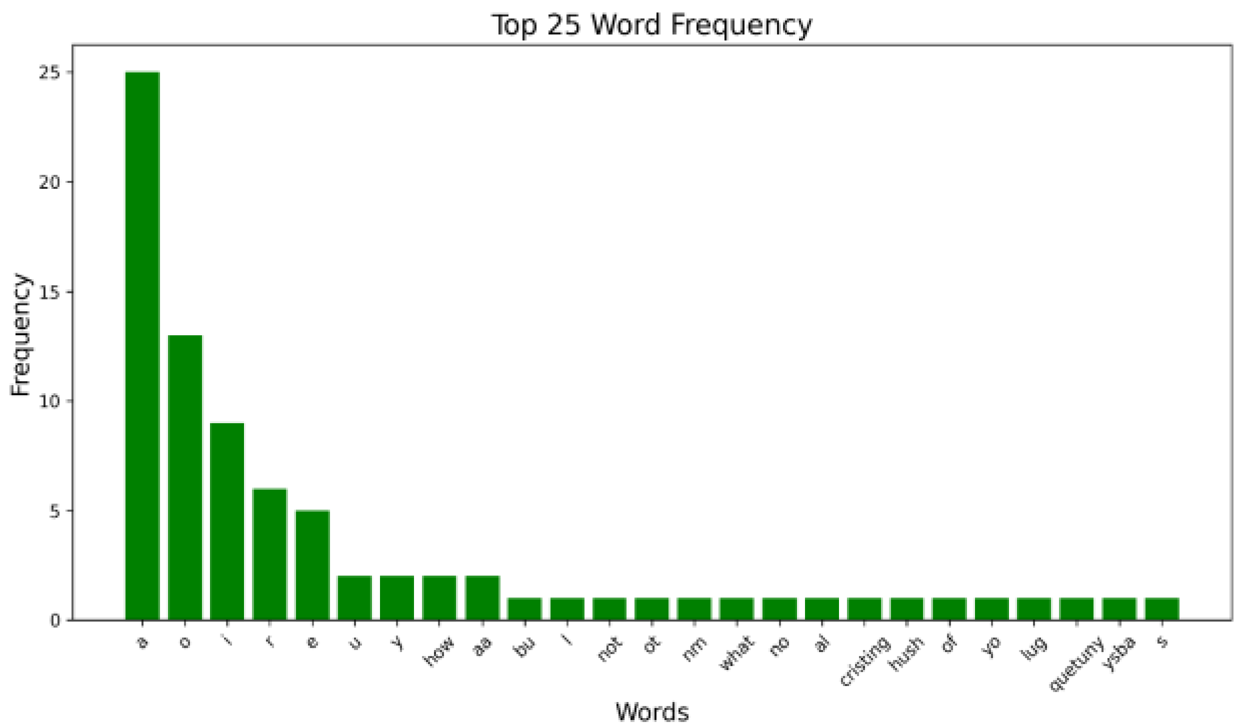

Figure 5.24 Word Frequency Distribution in Poultry Vocalizations presenting the top 25 words

The comparison of vowel frequency between healthy and unhealthy chickens (Figure 5.22) shows vowels "a", "o", and "e" being prevalent in both conditions, emphasizing the essential role of these sounds in poultry vocalization. Greater variation in vowel frequency, in unhealthy chickens, might suggest increased vocal effort (Figure 5.23) , perhaps indicating discomfort or a greater intent to promote signals of distress. These specific vowels that remained dominant in either healthy or unhealthy state may signify the importance of these sounds in the very framework of poultry communication. Unigram analysis (Figure 5.26) continues to indicate that the sound being the most common for both healthy and unhealthy chickens is "a", followed by "o", "i", and "e". Increased variety in unhealthy conditions points toward a diverse vocal effort, probably hinting at an increased need to express discomfort. Bigram analysis (Figure 5.27) further unveils that repeat phonemes such as "aa" are most dominant under distressing conditions, possibly signifying heightened attempts to communicate urgency or distress.

The word cloud (Figure 5.25) forms a visualization of vocal patterns between healthy and unhealthy chickens. For healthy chickens, mainly simple vowel-based sounds like "O", "AA", and "E" are used, signifying that they rely on simple vocal elements during routine communication. On the other side, sounds like "LUG", "BU", and "HUSH" are more prevalent in the case of the unhealthy, indicating an increased emission of context-dependent vocalizations related to stress or discomfort. The difference further shows how responsive vocal behavior in poultry is to shifting health conditions.

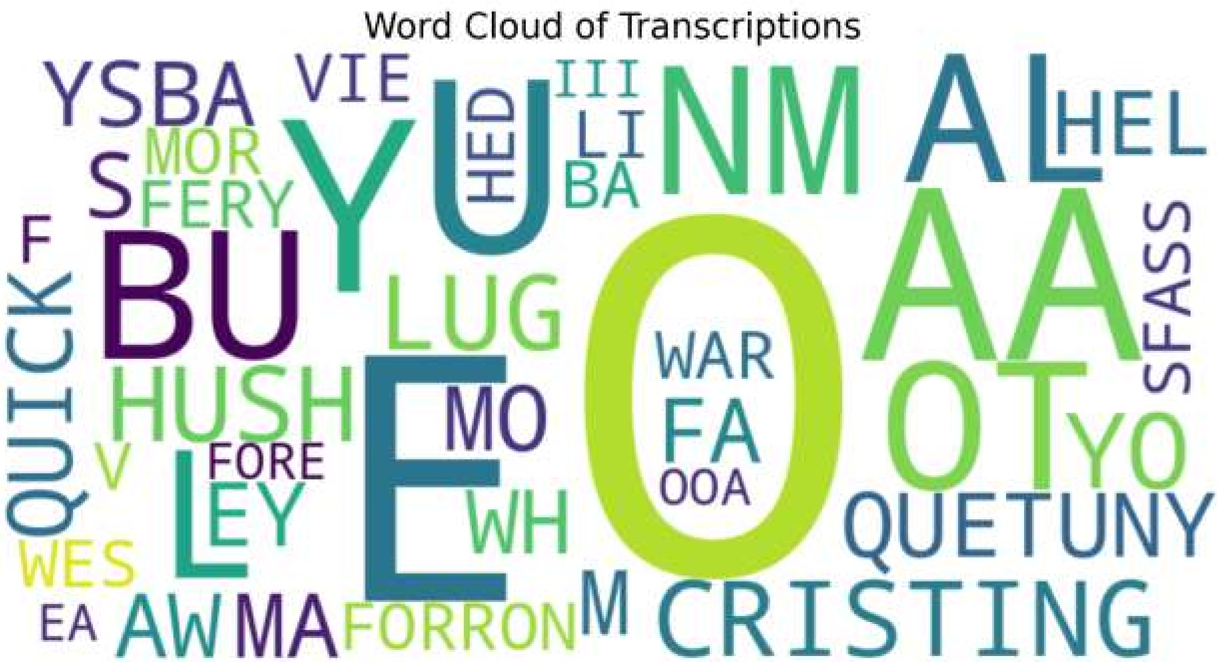

Figure 5.25. Word Cloud Analysis (Healthy vs. Unhealthy). Word cloud depicting vocal elements of healthy vs. unhealthy poultry, emphasizing differences in vocal patterns linked to health status.

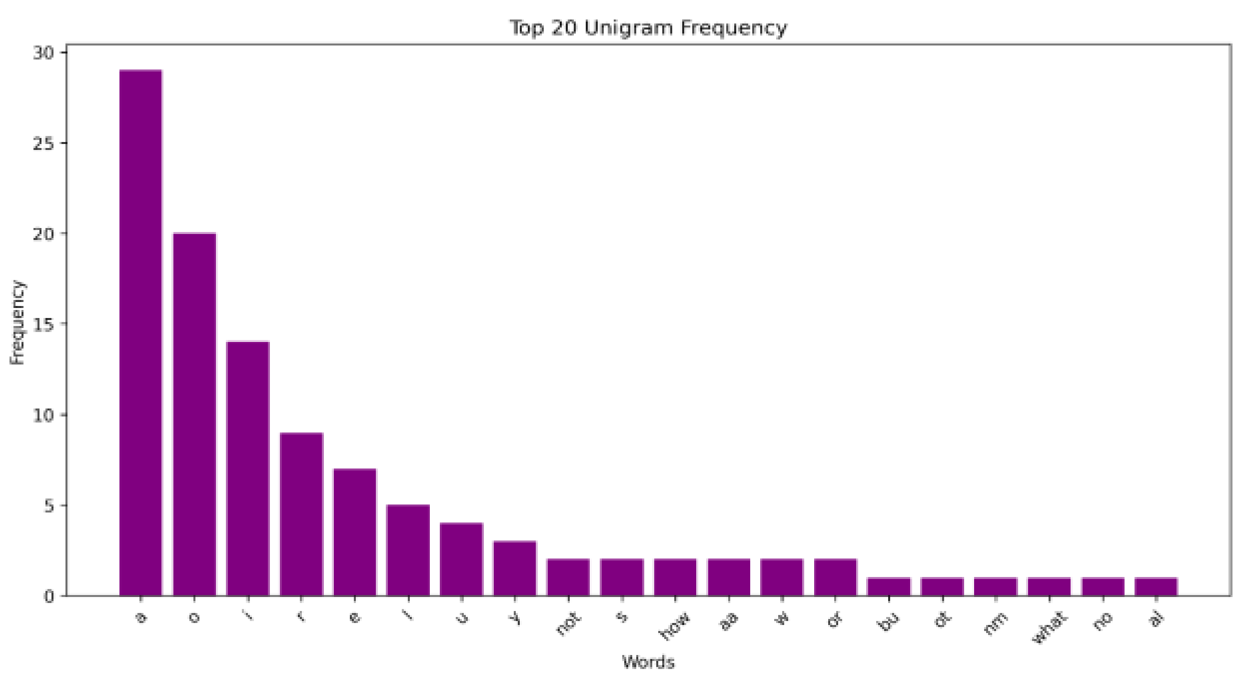

Figure 5.26 Unigram frequency distribution of healthy and unhealthy poultry vocalizations

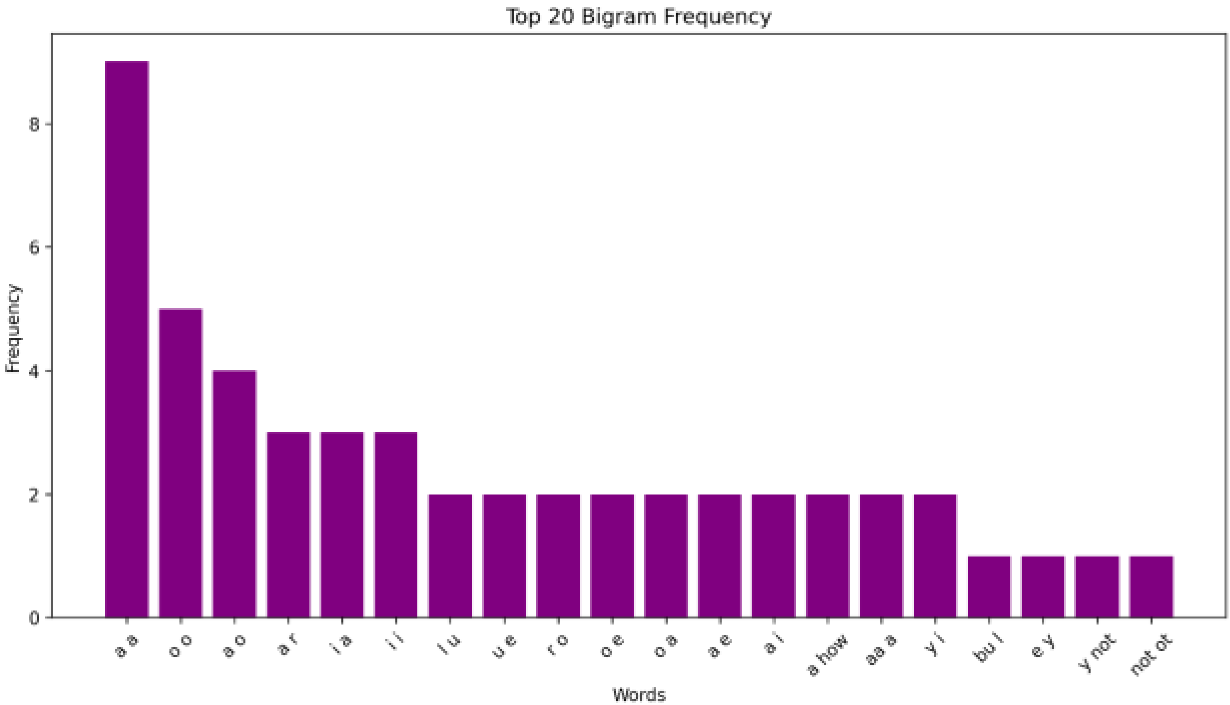

Figure 5.27 Bigram frequency distribution of healthy and unhealthy poultry vocalizations

5.2.8 Semantic Analysis of Chicken Language Dataset

With an analysis of the Chicken Language Dataset, similar sorts of results are found. The Chicken Language Dataset offers a very interesting perspective on how chickens interact among themselves. It was compiled in an attempt to bring out the diversity of chicken vocalization and language. The analysis opens significant patterns for vowel usage and frequency of various words and characters, including the overall complexity of linguistics involved in various illustrations contained in this work. The vowel-instance chart shows the emergence of certain vowels in the utterances made by the chickens.

The occurrence of the vowel "e" is the maximum, followed by "o" and "a," thus making them the notable triad in chicken communication. Vowels such as "i" and "u" are there, however, in relatively lesser occasions and may play a supporting role or be used for less frequent sounds. Imbalance in instance dominance of the two groups of "chicken language" could arise on account of some physiological influencing-limiting factors or the relevance of vowels in pronouncing types of emotional and environmental signals. The word frequency distribution (Figure 5.28) offers more detailed analysis revolving around the calls made by the chickens. Vowel words or calls "a," "o," "e" have once again been identified as being among the prime vocabulary of chickens, as is diagnosed in the chart of their vocalization.

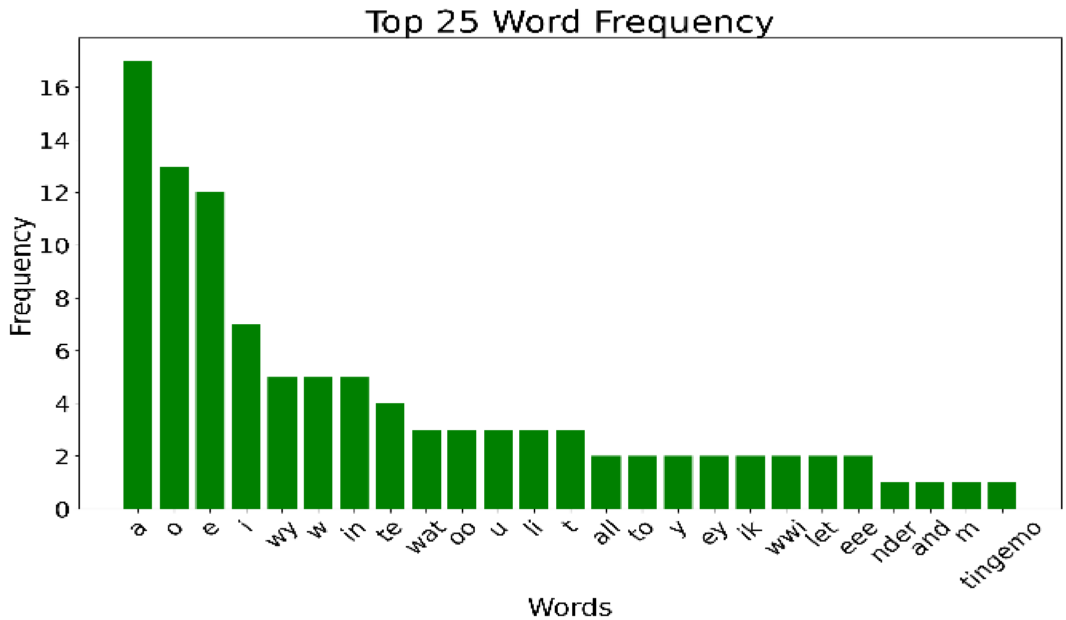

Figure 5.28 Top word Frequency in the Chicken Language Dataset.

Words such as "wy" or "in" in chicken communication might imply sequential vocalizations for alerting about food or danger. The large vocabulary might indicate that chickens have a blocked form of language organized to be used for different situations. The character frequency chart, on the other hand, tells us about language-making structures for the chicken calls. The letter e is reported as most frequent, having several occurrences of o, a, and t trailing it. These characters are basically shaping the calls of the chicken, thus consonant with vowels occupying the majority in this study's dataset.

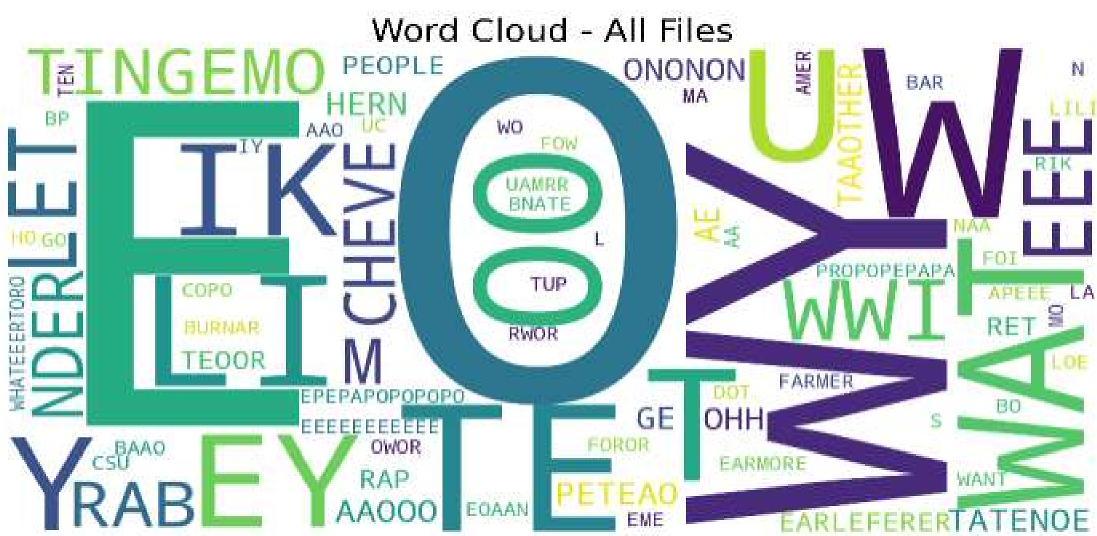

Figure 5.29 Word cloud Diagram of the Chicken Language dataset

The word cloud (Figure 5.29) gives vast information about the dataset and also the most relevant words that act according to those frequencies. Words like "T," "INGEMO," "EIK," and "WWI" are some of the significant ones exhibiting engagement patterns with some conceivable meaning for communication among chickens. The high frequency of vowels, the definite occurrence of the same words and characters in the middle of others, and the degree of variability in the word cloud give a probable verdict of a complex, robust system with situational application.

Such results emphasize the possibility of further development of audio analysis of the hen's calls, including training of artificial neural networks for automatic translation of hen's calls. By correlating specific calls with specific emotions or external influences, they can help improve animal welfare monitoring systems and deepen knowledge of communication systems in other species.

The sociological perspective provided by the study of the chicken language dataset is striking: There are multiplicity patterns of social and communication behavior amongst chickens. These varied sounds are used for the purpose of staying in one group or, and for some, the sound allows the chicken to find the other chickens who are somewhere in the area. Such sounds allow the chicken to feel a sense of safety within the group and maintain order. Some sounds either assist or direct which members of the group are to follow others, only further solidifying the group against possible external threats.

Also, we notice some differentiation in roles do take place amongst members of the in-group as vocalizations demonstrate this. These calls refer to actions that are different among roosters, chickens, and even chicks because they relate communication with their role in an appropriate manner. Also, the sounds emitted when chicks are the subject of guidance or mates are being called suggest a level of seniority and role differentiation in the group. Such differentiation maximises efficiency in social systems' operations and hence demonstrates an advanced type of social understanding in relation to roles. Another key aspect includes regard and sensitivity to emotions and situation. There are other sounds including the ones a chicken makes when they are hungry, thirsty, uncomfortable, or are feeling heat which promotes a need to bring the sound with the state they wish to project.

5.3 Comparative Analysis of Classical ML and Semantic Approaches

The two strands of analysis in this study—classical machine learning (based on structured acoustic features) and semantic approaches (based on transformer deep learning applied to raw or minimally preprocessed vocalizations)—together provide comprehensive insight into this issue of vocalization-dependent welfare assessment in poultry. The comparison goes into detail on how these methods perform relative to one another on different aspects of data representation and interpretability, being sensitive to and practically deploying under environmental conditions.

5.3.1 Feature Representation and Learning Dynamics

Classical ML models, particularly tree ensembles such as HistGradientBoosting and CatBoost, relied very much on explicit feature crafting using MFCC, spectral contrast, and ZCR. These models could, very much on a fine scale, distinguish the acoustic manifestations of stress or health status with accuracies as high as 99% (Dataset-1: health classification) and with a high degree of interpretability of feature importances.

On the contrary, semantic models based on wav2vec2.0 and BERT learned in an end-to-end fashion, modeling frequency contours, phonetic structures, and sentiment changes without any explicit feature crafting. They encoded perceptible shifts in pitch distributions and the emotional tone, for instance, a narrowed frequency range under stress (480-500 Hz) or vowel dominance after stress. Pre-trained language or audio encoder, upon which the semantic approach had to rely on allow generalization across very subtle signal patterns. However, it could not directly guarantee interpretability, as these internal representations or hidden embeddings do not actually map directly to physiological acoustic markers like MFCC_0 or spectral contrast.

5.3.2 Sensitivity to Dataset Size and Class Balance

Classical ML-based algorithms showed strength even though the setup was relatively modest, in terms of size, by efficiently applying tree regularization and boosting. However, the huge performance gap with TabNet (which clearly suffered from a lack of data) only supports the claim that deep tabular networks require much bigger sample size.

On the other hand, the semantic approach was able to take advantage of transfer learning based on pre-trained wav2vec2.0 and BERT, which provided an opportunity to learn efficiently even in cases where there is a paucity of labeled stress datasets. Fine-tuning required careful balancing to avoid overfitting, as we can observe a stable but slow loss convergence in BERT sentiment classification.

5.3.3 Interpretability, Physiological Relevance, and Consistency Across Contexts

One of the most significant aspects of comparison lies in how each methodology alludes to potential physiological and behavioral mechanisms on one hand and on the other hand, with what consistency they do so across different applications such as health detection, behavioral classification, and temporal stress adaptation. Classical methods of machine learning, especially tree-based ensemble methods such as HistGradientBoosting and CatBoost, were able to produce directly interpretable feature contributions. Subsequently, feature-importance rankings, together with correlation analyses, have shown that some acoustic features such as MFCC means, spectral contrast, and zero-crossing rate appeared to present a very strong biomarker for stress or illness and they were correlated to the variations in stress levels. Also, these features are well linked to physiological phenomena known to represent alterations of respiratory dynamics or the structure of the syrinx due to stress, thus providing excellent biological grounding for the patterns detected by the system. This bond inculcates confidence in the system and also paves the way for real-world applications where farmers or veterinarians may want to understand the perceived indicators for interventions.

Moreover, these classical models also maintained their performance through several experimental conditions involving health classification, behavioral call recognition, and preliminary temporal trend tracing. Their abilities to consistently select generally the same sets of discriminative features across tasks support their reliability in varying settings and further attest to the robustness of these handcrafted features as general acoustic indicators. The semantic transformer provided the complementary view. Being an acoustic-to-phonetic transformer, the wav2vec2.0 acoustic encoder along with BERT for sentiment and linguistic interpretation was less transparent to match with the actual physiological feature. Broader-level acoustic patterns with respect to the emotional and communicative aspects of poultry welfare were made latent. For instance, narrowing of pitch distributions and vowel dominance during post-stress phases, or more subtle changes in the intensities of sentiment from pre-stress balanced expressions to post-stress negatively skewed expressions, indicate the ways in which stress is manifested in the expressive vocal repertoire of the birds.

Also, our approaches are really useful when we track dynamic behavioural adaptations over time. This is observed in our temporal stress experiments where we can see how models capture the habituation is progressing and variability is also being sustained when under different stress conditions. Even when there is an absence of explicitly mapping of individual acoustic features like MFCC_7 or spectral contrast bands, it gives us an idea of how hens adapt their vocalizations to their environments. This in turn offers us insights of their physiology, emotion and social communication.

5.4 Towards a Hybrid Model for Poultry Welfare Monitoring

The discussions regarding the two complementary approaches raises another important possibility. We move towards a hybrid model that uses both of these approaches.

The approach can use the physiological interpretability and feature-level sensitivity of the classical machine learning approach with which it can trace specific bioacoustic markers like MFCC deviation or spectral contrast shifts that are linked to stress. Also, it can harness the contextual, emotional, and communicative depth that are captured by the semantic analysis done by the transformer models. We get deeper meanings such as the drift in sentiments, the rebalancing of phonetics and also the communication patterns that are adapted to stress.

5.4.1 Conceptual Design of the Hybrid Model

We envision the hybrid model to have a multi view input strategy. First we combine manually extracted acoustic features such as MFCCs, ZCRs , Spectral contrasts along with the raw waveform embeddings that we get from the wave2vec 2.0 encoder. This integration of two distinct feature representation will help the system to capture physiological markers that are well established with a more complex and holistic vocal signature.

The following architecture could also be a similar dual branch approach. The first branch is the classical feature encoder like gradient boosting model which can be optimized for the acoustic features that are structured. With the help of this branch, we can deliver interpretable results like feature importance and also the risk indicators. These help us link the direct physiological processes. The second branch would help us go deeper by incorporating a transformer based (BERT for example) that gives us the sentiments, complexity of the phonetics and also the temporal evolution of the vocal patterns influenced by various stressors. By using these two branches, the architecture can learn abstract, higher level contexts both from behaviour and emotional perspective from the birdsong.

The output from both these branches can be merged together into a new fusion layer which can concatenate their latent representations or extending further by using attention mechanisms that can be used to emphasize the most important information each branch. This fused representation in turn can be used feed an unified prediction layer that can simultaneously assess health status, stress levels, and broader behavioral contexts.

This method can be advantageous in multiple ways. The first and foremost advantage is the ability to achieve a higher robustness and generalization. This is possible because the tree based models stabilize learning on smaller dataset while the transformer pathway helps in uncovering the subtle sequential and emotional nuances that can often be missed by conventional features. More importantly, this combined approach can also help in cross validation. For example, Under stress, the pitch narrows and this can be corroborated with the acoustic markers and spectral shifts. This multi faceted approach not only improves the reliability of the model but also helps in deepening of our understanding of poultry welfare dynamics. This lays the foundation of a more precise, non invasive and an intelligent monitoring solution.

Chapter 6 – Conclusions, limitations and Future Directions

6.1 Leveraging Research Insights for Poultry Farming Management

The findings revealed by this exhaustive research into poultry vocalizations, including MFCCs, spectral descriptors, zero-crossing rates, feature importance maps, pitch distribution, phonetic composition, and sentiment dynamic analysis, pose opportunities that can be exploited for transformative poultry welfare monitoring and management in a large scale.

The detection of unique acoustic patterns caused by shifts in MFCC coefficients, drops in spectral contrast, pitch narrowing at the 480-500Hz band, and fluctuations in the distribution of emotional or phonetic content sets a solid platform on which welfare interventions in real time can be initiated. Relying on such signs, real-time alert can be sent to farmers upon detection of vocalizations triggered by stress or illness, thus allowing countermeasures to be put forth at an early stage. These interventions reduce outbreak risks while increasing general health conditions of the flocks, with mortality being lowered and productivity improved due to timely care.

We can also deduce how vocalizations features relate to the environmental factors. This opens up opprotunities for a more responsive, responsible farm management. For example, seeing stress-related calls from the perspective of spectral feature deviations or sentiment variations, or both, may indicate discomfort during high temperatures and hence could trigger an automated intervention in the ventilation or cooling systems to relieve heat stress. On the other hand, the knowledge as to how vocal patterns change under varying lighting regimes or with feed introduction delivers critical feedback for the optimization of both lighting schedules and feed formulations. Through these optimizations, feed efficiency and welfare are bettered and production outcomes gain in quality.

The analysis of vocalizations also reveals meaningful patterns that relate to the social and behavioral dynamics of poultry. Distinguishing call types associated with feeding, greetings, or distress-from both low-level acoustic cues and more phonetic or sentiment structural variants-brings an additional dimension of insight into how populations interact, thus enabling farmers to more strategically coordinate flock integration activities, thereby mitigating instances of aggressive behavior and associated injuries. Also, vocal cues related to mating behaviours can provide valuable insights for breeding programs. This increases the chance of a successful reproductive rates. By combining these vocal markers with other technologies for precision agriculture like weight monitoring or sensors for activity, we can yield predictive models that help us in anticipating the growth trends, social disruptions, thus guide us in targeting the required management actions.

The hybrid system of classical acoustic feature-based modeling coupled with an NLP-driven semantic pipeline can be directly used in commercial poultry farming. Using such a system would entail the constant evaluation of vocal streams across the farm environment with an emphasis on the identification of statistical anomalies in acoustic features as well as subtle changes in linguistic or sentiment structures derived from vocal transcriptions. Embedding such a hybrid pipeline into the day-to-day life of the farm would permit the buildup of intelligent dashboards presenting the welfare indicators in near real time, auto-adjusting climate controls or lighting systems from early signatures of stress and issuing operational alerts to direct on-site farm personnel toward inspections or intervention options. Over time, this hybrid approach may even learn the farm-specific vocal baseline and thus become increasingly sensitive to localized issues such as feed inconsistencies or breed-related stress responses. This fosters not only immediate welfare benefits, but also ties into long-term planning for biosecurity, optimized resource utilization, and farming practices that align better with sustainability and ethics.

6.2 Limitations and Future Directions

Classification of poultry vocalizations based on acoustic characteristics remains constrained by several key limitations. Although the datasets used represent different behavioral and physiological conditions, their overall scale and ecological diversity are relatively limited. Background noise, overlapping calls within group-housed settings, and class imbalance continue to pose challenges for robust model training and generalization, even after preprocessing and stratified sampling efforts. The lack of concurrent physiological or biochemical ground-truth data restricts the depth of biological interpretation, particularly when linking acoustic features to welfare-relevant internal states.

Future directions include the integration of multimodal sensor data to enrich contextual understanding. Combining acoustic features with environmental parameters such as temperature, humidity, and gas concentrations, as well as visual inputs related to movement and posture, could enable a more comprehensive welfare assessment framework. Advancements in edge computing and TinyML offer pathways for real-time, on-farm deployment of acoustic models using low-power, cost-effective devices—minimizing latency and reliance on cloud infrastructure. Incorporating self-supervised and few-shot learning approaches would further improve model adaptability in data-scarce conditions, especially for underrepresented behaviors or farm environments.

The wav2vec2.0 and BERT models in this study were originally pre-trained on large-scale human speech datasets. Although they were fine-tuned for poultry vocalizations, their underlying feature representations and language structures are themselves optimized for human phonetics and semantics. This mismatch may cause some difficulties for them to be sensitive to species-specific acoustic features or may give rise to suboptimal generalization while dealing with complex avian vocal behaviors.

Future work should be directed toward building or fine-tuning self-supervised models directly on large avian vocalization corpora, which could involve training wav2vec-type encoders from scratch on multi-species bird sound libraries or capitalizing on emerging bioacoustics-oriented models. Further, custom tokenizers and language models could be developed on bird calls that had been transcribed or clustered to provide a biologically meaningful embedding, which might then be integrated into detection pipelines requiring heightened sensitivity to subtle stress or welfare cues in poultry vocalizations.

Further refinement of this hybrid approach of classical acoustic models with NLP-driven semantic interpretations will open up enticing possibilities beyond proof-of-concept studies. Practical next steps could include farm-scale pilot deployments - where It detects abnormal events and reacts to them, changing ventilation, lighting, or feed dispenser settings based on an interpretation of stress cues in a closed-loop fashion. Also, cloud-connected dashboards may group vocal health cues from multiple farms to support data-driven decision-making at regional and national levels, enhancing and standardizing biosecurity and welfare criteria industry-wide. Over time, integrating such hybrid-computer-model constructs into operational standard farm protocols may lead to considering such systems as establishing a baseline for automated welfare audits and contributing to healthier flocks, more productive farming, and a more consumer-friendly production system in poultry.

In this study, the inter-model agreement was analyzed mainly in terms of Cohen's Kappa, a measure of how consistently two classifiers methodically label the same set of samples under the same input parameters. This agreement allows us to verify if ensemble learning or model stacking makes sense from an inter-model agreement perspective. Since every model was given the same label set and training partitions, this also allows for a fair intra-model Kappa comparison.

As a more classical use example, a Kappa index can also be computed between model predictions and human annotations. In subsequent studies, if expert annotators begin labeling vocalizations manually (such as distress, contact, alarm), then Kappa statistics could be extended to human versus model agreement, which would allow for more thorough assessment of model interpretability and operational reliability.

Long-term implementation of AI welfare monitoring systems will always consider behavioral plasticity — the possibility for animals to adapt their behavior when confronted with prolonged monitoring conditions. In the case of the vocalization-based models, the question remains whether chickens may alter their vocal behavior in a way that either manipulates or circumvents the system. Once the model triggers some sort of environmental modification (including changes in light, temperature, or human intervention) based on vocal cues, changes could occur in how the sounds/emission patterns of the birds are associated with consequences. At present, we have no observation in our study about such feedback-loop-induced vocal adaptation, but it certainly remains an important topic to be studied over a longitudinal timeline in real-world situations.

Furthermore, the vocal behavior of chickens is not purely instinctual but shaped by early experience, environmental enrichment, and social imprinting. Chicks that have experienced maternal vocalizations either during incubation or in early rearing differ in their acoustic repertoire from those raised either in isolation or in artificial incubators. These contextual and developmental influences can add a layer of complexity to vocal behavior that cannot fully be accounted for in models trained only on treatment-based labels (stress versus no-stress). Factors such as peer-group interactions, housing conditions, previous handling experiences, and even cross-species acoustic exposure (for example, exposure to sounds produced by dogs, humans, or machinery) may be influencing vocal output and stress physiology.

This thesis attempted to maintain environmental consistency within the datasets used, but the absence of metadata on rearing practices, maternal exposure, and flock dynamics significantly limited the model's ability to generalize over highly variant farming scenarios. Hence, it is recommended that future work looks to combine behavioral ethology and AI, and integrate metadata pertaining to social context, rearing conditions, and environmental history. This may facilitate the development of more robust and interpretable models that incorporate individual learned vocal behaviors, as well as those that are socially conditioned-a very important aspect when trying to address flock-wide welfare assessments in real life.

References

1. Serbessa TA, Geleta YG, Terfa IO. Review on diseases and health management of poultry and swine. Int J Avian & Wildlife Biol. 2023;7(1):27-38. DOI: https://doi.org/10.15406/ijawb.2023.07.00187
2. Lawal, R.A., Martin, S.H., Vanmechelen, K. et al. The wild species genome ancestry of domestic chickens. BMC Biol 18, 13 (2020). https://doi.org/10.1186/s12915-020-0738-1
3. He, P.; Chen, Z.; Yu, H.; Hayat, K.; He, Y.; Pan, J.; Lin, H. Research Progress in the Early Warning of Chicken Diseases by Monitoring Clinical Symptoms. Appl. Sci. 2022, 12, 5601. https://doi.org/10.3390/app12115601
4. 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, 130, 79–96. DOI: https://doi.org/10.1016/j.anbehav.2017.05.025
5. 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
6. Ungar, E.D.; Nevo, Y. Rhythms, Patterns and Styles in the Jaw Movement Activity of Beef Cattle on Rangeland as Revealed by Acoustic Monitoring. Sensors 2025, 25, 1210. https://doi.org/10.3390/s25041210
7. Gao, G.; Ma, Y.; Wang, J.; Li, Z.; Wang, Y.; Bai, H. CFR-YOLO: A Novel Cow Face Detection Network Based on YOLOv7 Improvement. Sensors 2025, 25, 1084. https://doi.org/10.3390/s25041084
8. Dewmini, H.; Meedeniya, D.; Perera, C. Elephant Sound Classification Using Deep Learning Optimization. Sensors 2025, 25, 352. https://doi.org/10.3390/s25020352
9. Karaaslan, M.; Turkoglu, B.; Kaya, E.; Asuroglu, T. Voice Analysis in Dogs with Deep Learning: Development of a Fully Automatic Voice Analysis System for Bioacoustics Studies. Sensors 2024, 24, 7978. https://doi.org/10.3390/s24247978
10. Brennan, S.E.; et al. The PRISMA 2020 statement: An updated guideline for reporting systematic reviews. BMJ (Clin. Res. ed.) 2021, 372, n71. https://doi.org/10.1136/bmj.n71.
11. Soster, P.; Grzywalski, T.; Hou, Y.; Thomas, P.; Dedeurwaerder, A.; De Gussen, M.; Tuyttens, F.; Devos, P.; Botteldooren, D.; Antonissen, G. A machine learning approach for broiler chicken vocalization monitoring. SSRN Electron. J. 2024. https://doi.org/10.2139/ssrn.4999753
12. Michaud, F.; Sueur, J.; Le Cesne, M.; Haupert, S. Unsupervised classification to improve the quality of a bird song recording dataset. Ecol. Inform. 2023, 74, 101952. https://doi.org/10.1016/j.ecoinf.2022.101952

13. Bravo Sanchez, F.J.; Hossain, R.; English, N.B.; Moore, S.T. Bioacoustic classification of avian calls from raw sound waveforms with an open-source deep learning architecture. Sci. Rep. 2021, 11, 15733. https://doi.org/10.1038/s41598-021-95076-6
14. Umarani, M.; Meyyappan, S.; Vallathan, G.; Karthi, G. LSTM-based vocalization analysis for identification and classification of avian acoustics. In Proceedings of the International Conference on Computational Intelligence for Green and Sustainable Technologies (ICCIGST), Vijayawada, India, 18-19 July 2024. https://doi.org/10.1109/ICCIGST60741.2024.10717534.
15. Pereira, E.; Nääs, I.d.A.; Ivale, A.H.; Garcia, R.G.; Lima, N.D.d.S.; Pereira, D.F. Energy assessment from broiler chicks' vocalization might help improve welfare and production. Animals 2023, 13, 15. https://doi.org/10.3390/ani13010015.
16. 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. Classification of vocalization recordings of laying hens and cattle using convolutional neural network models. J. Biosyst. Eng. 2021, 46, 217–224. https://doi.org/10.1007/s42853-021-00101-1.
17. Thomas, P.; Grzywalski, T.; Hou, Y.; de Carvalho, P.S.; De Gussem, M.; Antonissen, G.; Botteldooren, D. Using a neural network-based vocalization detector for broiler welfare monitoring. In Proceedings of the 10th Convention of the European Acoustics Association, Turin, Italy, 11–15 September 2023. https://doi.org/10.61782/fa.2023.0474.
18. Prabakaran, D.; Sriuppili, S. Speech processing: MFCC based feature extraction techniques—An investigation. J. Phys. Conf. Ser. 2021, 1717, 012009. https://doi.org/10.1088/1742-6596/1717/1/012009.
19. Davis, S.; Mermelstein, P. Comparison of parametric representations for monosyllabic word recognition in continuously spoken sentences. IEEE Trans. Acoust. Speech Signal Process. 1980, 28, 357–366. https://doi.org/10.1109/TASSP.1980.1163420.
20. Sattar, F. A context-aware method-based cattle vocal classification for livestock monitoring in smart farm. Chem. Proc. 2022, 10, 89. https://doi.org/10.3390/IOCAG2022-12233.
21. Puswal, S.M.; Liang, W. Acoustic features and morphological parameters of the domestic chickens. Poult. Sci. 2024, 103, 103758. https://doi.org/10.1016/j.psj.2024.103758.
22. Zhong, M.; Taylor, R.; Bates, N.; Christey, D.; Basnet, H.; Flippin, J.; Lavista Ferres, J. Acoustic detection of regionally rare bird species through deep convolutional neural networks. Ecol. Inform. 2021, 64, 101333. https://doi.org/10.1016/j.ecoinf.2021.101333.
23. Henri, E.J.; Mungloo-Dilmohamud, Z. A deep transfer learning model for the identification of bird songs: A case study for Mauritius. In Proceedings of the International Conference on Electrical, Computer, Communications and Mechatronics

Engineering (ICECCME), Mauritius, 7–8 October 2021. https://doi.org/10.1109/ICECCME52200.2021.9590917.
24. Romero-Mujalli, D.; Bergmann, T.; Zimmermann, A.; Scheumann, M. Utilizing DeepSqueak for automatic detection and classification of mammalian vocalizations: A case study on primate vocalizations. Sci. Rep. 2021, 11, 24463. https://doi.org/10.1038/s41598-021-03941-1
25. Thomas, M.; Jensen, F.H.; Averly, B.; Demartsev, V.; Manser, M.B.; Sainburg, T.; Roch, M.A. A practical guide for generating unsupervised, spectrogram-based latent space representations of animal vocalizations. J. Anim. Ecol. 2022, 91, 1567–1581. https://doi.org/10.1111/1365-2656.13754.
26. Mao, A.; Giraudet, C.S.E.; Liu, K.; Nolasco, I.D.A.; Xie, Z.; Xie, Z.; McElligott, A.G. Automated identification of chicken distress vocalizations using deep learning models. J. R. Soc. Interface 2022, 19, 20210921. https://doi.org/10.1098/rsif.2021.0921.
27. Mangalam, K.; Sarkar, S.; Dogra, Y.; Saini, M.; Goel, N. What Did the Chicken Say: A Multi-class Classification Method on Chicken Vocalizations. In Proceedings of the International Conference on Systems and Technologies for Smart Agriculture, Kolkata, India, 19–20 December 2023; Saha, H.N., Ray, H., Bradford, P.G., Eds.; Springer: Singapore, 2025; pp. 667–676. https://doi.org/10.1007/978-981-97-5157-0_55.
28. Li, Z.; Zhang, T.; Cuan, K.; Fang, C.; Zhao, H.; Guan, C.; Yang, Q.; Qu, H. Sex detection of chicks based on audio technology and deep learning methods. Animals 2022, 12, 3106. https://doi.org/10.3390/ani12223106.
29. Neethirajan, S. Vocalization patterns in laying hens—An analysis of stress-induced audio responses. bioRxiv 2023. https://doi.org/10.1101/2023.12.26.573338.
30. Sainburg, T.; Thielk, M.; Gentner, T.Q. Finding, visualizing, and quantifying latent structure across diverse animal vocal repertoires. PLoS Comput. Biol. 2020, 16, e1008228. https://doi.org/10.1371/journal.pcbi.1008228.
31. Ginovart-Panisello, G.J.; Iriondo, I.; Panisello Monjo, T.; Riva, S.; Garcia, R.; Valls, J.; Alsina-Pagès, R.M. Acoustic detection of the effects of prolonged fasting on newly hatched broiler chickens. Comput. Electron. Agric. 2024, 219, 108763. https://doi.org/10.1016/j.compag.2024.108763
32. Tao, W.; Wang, G.; Sun, Z.; Xiao, S.; Wu, Q.; Zhang, M. Recognition method for broiler sound signals based on multi-domain sound features and classification model. Sensors 2022, 22, 7935. https://doi.org/10.3390/s22207935.
33. Bermant, P.C.; Bronstein, M.M.; Wood, R.J.; Gero, S.; Gruber, D.F. Deep machine learning techniques for the detection and classification of sperm whale bioacoustics. Sci. Rep. 2019, 9, 12588. https://doi.org/10.1038/s41598-019-48909-4.
34. Terasaka, D.T.; Martins, L.E.; dos Santos, V.A.; Ventura, T.M.; de Oliveira, A.G.; Pedroso, G.S.G. Bird audio segmentation: Audio segmentation to build bird training datasets. In Proceedings of the Anais Estendidos do Workshop de Computação

Aplicada à Gestão do Meio Ambiente e Recursos Naturais (WCAMA), Brasília, Brazil, 21–25 July 2024; pp. 22–29. https://doi.org/10.5753/wcama.2024.2055.
35. Merino Recalde, N. pykanto: A python library to accelerate research on wild bird song. Methods Ecol. Evol. 2023, 14, 1994–2002. https://doi.org/10.1111/2041-210X.14155.
36. Swaminathan, B.; Jagadeesh, M.; Subramaniyaswamy, V. Multi-label classification for acoustic bird species detection using transfer learning approach. Ecol. Inform. 2024, 80, 102471. https://doi.org/10.1016/j.ecoinf.2024.102471.
37. Du, X.; Carpentier, L.; Teng, G.; Liu, M.; Wang, C.; Norton, T. Assessment of laying hens' thermal comfort using sound technology. Sensors 2020, 20, 473. https://doi.org/10.3390/s20020473.
38. Bhandekar, A.; Udutalapally, V.; Das, D. Acoustic based chicken health monitoring in smart poultry farms. In Proceedings of the 2023 IEEE International Symposium on Smart Electronic Systems (iSES), Ahmedabad, India, 18–20 December 2023; pp. 224–229. https://doi.org/10.1109/iSES58672.2023.00054.
39. Ginovart-Panisello, G.J.; Iriondo, I.; Panisello Monjo, T.; Riva, S.; Cancer, J.C.; Alsina-Pagès, R.M. Acoustic detection of vaccine reactions in hens for assessing anti-inflammatory product efficacy. Appl. Sci. 2024, 14, 2156. https://doi.org/10.3390/app14052156.
40. Cuan, K.; Zhang, T.; Li, Z.; Huang, J.; Ding, Y.; Fang, C. Automatic Newcastle disease detection using sound technology and deep learning method. Comput. Electron. Agric. 2022, 194, 106740. https://doi.org/10.1016/j.compag.2022.106740.
41. Cuan, K.; Zhang, T.; Huang, J.; Fang, C.; Guan, Y. Detection of avian influenza-infected chickens based on a chicken sound convolutional neural network. Comput. Electron. Agric. 2020, 178, 105688. https://doi.org/10.1016/j.compag.2020.105688.
42. Xu, R.-Y.; Chang, C.-L. Deep Learning-Based Poultry Health Diagnosis: Detecting Abnormal Feces and Analyzing Vocalizations. In Proceedings of the 2024 10th International Conference on Applied System Innovation (ICASI), Kyoto, Japan, 17–21 April 2024; pp. 55–57. https://doi.org/10.1109/ICASI60819.2024.10547723.
43. Gupta, G.; Kshirsagar, M.; Zhong, M.; Gholami, S.; Lavista Ferres, J. Comparing recurrent convolutional neural networks for large scale bird species classification. Sci. Rep. 2021, 11, 17085. https://doi.org/10.1038/s41598-021-96446-w.
44. Huang, J.; Zhang, T.; Cuan, K.; Fang, C. An intelligent method for detecting poultry eating behaviour based on vocalization signals. Comput. Electron. Agric. 2021, 180, 105884. https://doi.org/10.1016/j.compag.2020.105884.
45. Hassan, E.; Elbedwehy, S.; Shams, M.Y.; ElHafeez, T.A.; ElRashidy, N. Optimizing poultry audio signal classification with deep learning and burn layer fusion. J. Big Data 2024, 11, 135. https://doi.org/10.1186/s40537-024-00985-8.
46. Laleye, F.A.A.; Mousse, M.A. Attention-based recurrent neural network for automatic behavior laying hen recognition. Multimed. Tools Appl. 2024, 83, 62443–62458. https://doi.org/10.1007/s11042-024-18241-9.

47. Hu, S.; Chu, Y.; Wen, Z.; Zhou, G.; Sun, Y.; Chen, A. Deep learning bird song recognition based on MFF-ScSEnet. Ecol. Indic. 2023, 154, 110844. https://doi.org/10.1016/j.ecolind.2023.110844.
48. Ghani, B.; Kalkman, V.J.; Planqué, B.; Vellinga, W.-P.; Gill, L.; Stowell, D. Generalization in birdsong classification: Impact of transfer learning methods and dataset characteristics. Front. Bird Sci. 2024, 3, 1515383. https://doi.org/10.48550/arXiv.2409.15383.
49. Bianco, M.J.; Gerstoft, P.; Traer, J.; Ozanich, E.; Roch, M.A.; Gannot, S.; Deledalle, C.-A. Machine learning in acoustics: Theory and applications. J. Acoust. Soc. Am. 2019, 146, 3590–3628. https://doi.org/10.1121/1.5133944.
50. McGinn, K.; Kahl, S.; Peery, M.Z.; Klinck, H.; Wood, C.M. Feature embeddings from the BirdNET algorithm provide insights into avian ecology. Ecol. Inform. 2023, 74, 101995. https://doi.org/10.1016/j.ecoinf.2023.101995.
51. Vaswani, A.; Shazeer, N.; Parmar, N.; Uszkoreit, J.; Jones, L.; Gomez, A.N.; Kaiser, Ł.; Polosukhin, I. Attention is all you need. arXiv 2017, arXiv:1706.03762. https://doi.org/10.48550/arXiv.1706.03762.
52. Menezes, G.L.; Mazon, G.; Ferreira, R.E.P.; Cabrera, V.E.; Dorea, J.R.R. Artificial intelligence for livestock: a narrative review of the applications of computer vision systems and large language models for animal farming. Anim. Front. 2024, 14, 42–53. https://doi.org/10.1093/af/vfae048.
53. Devlin, J.; Chang, M.-W.; Lee, K.; Toutanova, K. BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. In 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); Association for Computational Linguistics: Stroudsburg, PA, USA, 2019; pp. 4171–4186. https://doi.org/10.18653/v1/N19-1423.
54. Abzaliev, A.; Pérez Espinosa, H.; Mihalcea, R. Towards Dog Bark Decoding: Leveraging Human Speech Processing for Automated Bark Classification. arXiv 2024, arXiv:2404.18739. https://doi.org/10.48550/arXiv.2404.18739.
55. Sarkar, E.; Magimai-Doss, M. Comparing Self-Supervised Learning Models Pre-Trained on Human Speech and Animal Vocalizations for Bioacoustics Processing. arXiv 2025, arXiv:2501.05987. https://doi.org/10.48550/arXiv.2501.05987.
56. Neethirajan, S. Adapting a Large-Scale Transformer Model to Decode Chicken Vocalizations: A Non-Invasive AI Approach to Poultry Welfare. AI 2025, 6, 65. https://doi.org/10.3390/ai6040065.
57. Morita, T.; Koda, H.; Okanoya, K.; Tachibana, R.O. Measuring context dependency in birdsong using artificial neural networks. PLoS Comput. Biol. 2021, 17, e1009707. https://doi.org/10.1371/journal.pcbi.1009707.
58. Gong, Y.; Chung, Y.-A.; Glass, J. AST: Audio Spectrogram Transformer. arXiv 2021, arXiv:2104.01778. https://doi.org/10.48550/arXiv.2104.01778.

59. Baevski, A.; Zhou, H.; Mohamed, A.; Auli, M. wav2vec 2.0: A Framework for Self-Supervised Learning of Speech Representations. arXiv 2020, arXiv:2006.11477. https://doi.org/10.48550/arXiv.2006.11477.
60. Wang, T.S.; Li, X.; Zhang, C.; Wu, M.; Zhu, K.Q. Phonetic and Lexical Discovery of Canine Vocalization. In Findings of the Association for Computational Linguistics: EMNLP 2024; Association for Computational Linguistics: Miami, FL, USA, 2024; pp. 13972–13983. https://doi.org/10.18653/v1/2024.findings-emnlp.816.
61. Mørk, J.; Bovbjerg, H.S.; Kiss, G.; Tan, Z.-H. Noise-robust keyword spotting through self-supervised pretraining. arXiv 2024, arXiv:2403.18560. https://doi.org/10.48550/arXiv.2403.18560.
62. Brydinskyi, V.; Sabodashko, D.; Khoma, Y.; Podpora, M.; Konovalov, A.; Khoma, V. Enhancing automatic speech recognition with personalized models: Improving accuracy through individualized fine-tuning. IEEE Access 2024, 12, 116649–116656. https://doi.org/10.1109/ACCESS.2024.3443811.
63. Tosato, G.; Shehata, A.; Janssen, J.; Kamp, K.; Jati, P.; Stowell, D. Auto deep learning for bioacoustic signals. arXiv 2023, arXiv:2311.04945v2. https://arxiv.org/abs/2311.04945.
64. Collias, N.; Joos, M. The spectrographic analysis of sound signals of the domestic fowl. Behaviour 1953, 5, 175–188. Available online: https://www.jstor.org/stable/4532776 (27 June 2025).
65. van den Heuvel, H.; Youssef, A.; Grat, L.M.; Neethirajan, S. Quantifying the effect of an acute stressor in laying hens using thermographic imaging and vocalisations. bioRxiv 2022. https://doi.org/10.1101/2022.07.31.502171.
66. Collins, S.A.; Herborn, K.; Sufka, K.J.; Asher, L.; Brilot, B. Do I sound anxious? Emotional arousal is linked to changes in vocalisations in domestic chicks (Gallus gallus dom.). Appl. Anim. Behav. Sci. 2024, 277, 106359. https://doi.org/10.1016/j.applanim.2024.106359.
67. Lev-Ron, T.; Yitzhaky, Y.; Halachmi, I.; Druyan, S. Classifying vocal responses of broilers to environmental stressors via artificial neural network. Animal 2025, 19, 101378. https://doi.org/10.1016/j.animal.2024.101378.
68. Zhao, S.; Cui, W.; Yin, G.; Wei, H.; Li, J.; Bao, J. Effects of different auditory environments on behavior, learning ability, and fearfulness in 4-week-old laying hen chicks. Animals 2023, 13, 3022. https://doi.org/10.3390/ani13193022.
69. Edgar, J.L.; Lowe, J.C.; Paul, E.S.; Nicol, C.J. Avian maternal response to chick distress. Proc. R. Soc. B Biol. Sci. 2011, 278, 3129–3134. https://doi.org/10.1098/rspb.2010.2701.
70. Zimmerman, P.H. “Say what?” Vocalisation as an Indicator of Welfare in the Domestic Laying Hen. Ph.D Thesis, Wageningen University, Wageningen, The Netherlands, 1999; ISBN 90-5808-159-1.

71. McGrath, N.; Phillips, C.J.C.; Burman, O.H.P.; Dwyer, C.M.; Henning, J. Humans can identify reward-related call types of chickens. R. Soc. Open Sci. 2024, 11, 231284. https://doi.org/10.1098/rsos.231284.
72. Neethirajan, S. Decoding the Language of Chickens—An Innovative NLP Approach to Enhance Poultry Welfare. biorXiv 2024. https://doi.org/10.1101/2024.04.29.591707.
73. 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), Glasgow, UK, 2–5 December 2024. Association for Computing Machinery: New York, NY, USA. https://doi.org/10.1145/3702336.3702344.
74. Schober, J.M.; Merritt, J.; Ulrey, M.; Yap, T.Y.; Lucas, J.R.; Fraley, G.S. Vocalizations of the Pekin duck (Anas platyrhynchos domesticus): How stimuli, sex, and social groups affect their vocal repertoire. Poult. Sci. 2024, 103, 103738. https://doi.org/10.1016/j.psj.2024.103738.
75. Neethirajan, S. From predictive analytics to emotional recognition—The evolving landscape of cognitive computing in animal welfare. Int. J. Cogn. Comput. Eng. 2024, 5, 123–131. https://doi.org/10.1016/j.ijcce.2024.02.003.
76. Cai, J.; Yan, Y.; Cheok, A. Deciphering Avian Emotions: A Novel AI and Machine Learning Approach to Understanding Chicken Vocalizations. Res. Sq. 2023. https://doi.org/10.21203/rs.3.rs-3034567/v1.
77. Gavojdian, D.; Mincu, M.; Lazebnik, T.; Oren, A.; Nicolae, I.; Zamansky, A. BovineTalk: Machine learning for vocalization analysis of dairy cattle under the negative affective state of isolation. Front. Vet. Sci. 2024, 11, 1357109. https://doi.org/10.3389/fvets.2024.1357109.
78. Lavner, Y.; Pérez-Granados, C. Editorial: Computational bioacoustics and automated recognition of bird vocalizations: New tools, applications and methods for bird monitoring. Front. Bird Sci. 2024, 3, 1518077. https://doi.org/10.3389/fbirs.2024.1518077.
79. Fontana, I.; Tullo, E.; Butterworth, A.; Guarino, M. An innovative approach to predict the growth in intensive poultry farming. Comput. Electron. Agric. 2015, 119, 178–183. https://doi.org/10.1016/j.compag.2015.10.001.
80. Karatsiolis, S.; Panagi, P.; Vassiliades, V.; Kamilaris, A.; Nicolaou, N.; Stavrakis, E. Towards understanding animal welfare by observing collective flock behaviors via AI-powered analytics. Ann. Comput. Sci. Inf. Syst. 2024, 39, 643–648. https://doi.org/10.15439/2024F2064.
81. Manteuffel, G.; Puppe, B.; Schön, P.C. Vocalization of farm animals as a measure of welfare. Appl. Anim. Behav. Sci. 2004, 88, 163–182. https://doi.org/10.1016/j.applanim.2004.02.012.
82. Güntürkün, O. The avian 'prefrontal cortex' and cognition. Curr. Opin. Neurobiol. 2005, 15, 686–693. https://doi.org/10.1016/j.conb.2005.10.003.

83. Galef, B.G.; Laland, K.N. Social learning in animals: Empirical studies and theoretical models. BioScience 2005, 55, 489–499. https://doi.org/10.1641/0006-3568(2005)055[0489:SLIAES]2.0.CO;2.
84. Rugani, R.; Fontanari, L.; Simoni, E.; Regolin, L.; Vallortigara, G. Arithmetic in newborn chicks. Proc. R. Soc. B Biol. Sci. 2009, 276, 2451–2460. https://doi.org/10.1098/rspb.2009.0044.
85. Mellor, D.J.; Beausoleil, N.J.; Littlewood, K.E.; McLean, A.N.; McGreevy, P.D.; Jones, B.; Wilkins, C. The 2020 Five Domains Model: Including Human-Animal Interactions in Assessments of Animal Welfare. Animals 2020, 10, 1870. https://doi.org/10.3390/ani10101870.
86. Herrando, C.; Constantinides, E. Emotional contagion: A brief overview and future directions. Front. Psychol. 2021, 12, 712606. https://doi.org/10.3389/fpsyg.2021.712606.
87. 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. Enhancing poultry health management through machine learning-based analysis of vocalization signals dataset. Data Brief 2023, 50, 109528. https://doi.org/10.1016/j.dib.2023.109528.
88. Amirivojdan, A.; Nasiri, A.; Zhou, S.; Zhao, Y.; Gan, H. ChickenSense: A Low-Cost Deep Learning-Based Solution for Poultry Feed Consumption Monitoring Using Sound Technology. AgriEngineering 2024, 6, 2115–2129. https://doi.org/10.3390/agriengineering6030124.
89. Srinivasagan, R.; El Sayed, M.S.; Al-Rasheed, M.I.; Alzahrani, A.S. Edge intelligence for poultry welfare: Utilizing tiny machine learning neural network processors for vocalization analysis. PLoS ONE 2025, 20, e0316920. https://doi.org/10.1371/journal.pone.0316920.
90. Gibb, R.; Browning, E.; Glover-Kapfer, P.; Jones, K.E. Emerging opportunities and challenges for passive acoustics in ecological assessment and monitoring. Methods Ecol. Evol. 2019, 10, 169–185. https://doi.org/10.1111/2041-210X.13101.
91. Schneider, S.; Hammerschmidt, K.; Dierkes, P.W. Introducing the Software CASE (Cluster and Analyze Sound Events) by Comparing Different Clustering Methods and Audio Transformation Techniques Using Animal Vocalizations. Animals 2022, 12, 2020. https://doi.org/10.3390/ani12162020.
92. Nicholson, D. Crowsetta: A Python tool to work with any format for annotating animal vocalizations and bioacoustics data. J. Open Source Softw. 2023, 8, 5338. https://doi.org/10.21105/joss.05338.
93. Lapp, S.; Rhinehart, T.; Freeland-Haynes, L.; Khilnani, J.; Syunkova, A.; Kitzes, J. OpenSoundscape: An open-source bioacoustics analysis package for Python. Methods Ecol. Evol. 2023, 14, 2321–2328. https://doi.org/10.1111/2041-210X.14196.
94. Rauch, L.; Schwinger, R.; Wirth, M.; Heinrich, R.; Huseljic, D.; Herde, M.; Lange, J.; Kahl, S.; Sick, B.; Tomforde, S.; et al. BirdSet: A Large-Scale Dataset for Audio

Classification in Avian Bioacoustics. arXiv 2025, arXiv:2403.10380. https://doi.org/10.48550/arXiv.2403.10380.
95. Sasek, J.; Allison, B.; Contina, A.; Knobles, D.; Wilson, P.; Keitt, T. Semiautomated generation of species-specific training data from large, unlabeled acoustic datasets for deep supervised birdsong isolation. PeerJ 2024, 12, e17854. https://doi.org/10.7717/peerj.17854.
96. Ranjard, L.; Ross, H.A. Unsupervised bird song syllable classification using evolving neural networks. J. Acoust. Soc. Am. 2008, 123, 4358–4368. https://doi.org/10.1121/1.2903861.
97. Cohen, Y.; Nicholson, D.A.; Sanchioni, A.; Mallaber, E.K.; Skidanova, V.; Gardner, T.J. Automated annotation of birdsong with a neural network that segments spectrograms. eLife 2022, 11, e63853. https://doi.org/10.7554/eLife.63853.
98. Sethi, S.S.; Bick, A.; Chen, M.-Y.; Crouzeilles, R.; Hillier, B.V.; Lawson, J.; Banks-Leite, C. Large-scale avian vocalization detection delivers reliable global biodiversity insights. Proc. Natl. Acad. Sci. USA 2024, 121, e2315933121. https://doi.org/10.1073/pnas.2315933121.
99. Lostanlen, V.; Cramer, A.; Salamon, J.; Farnsworth, A.; Van Doren, B.M.; Kelling, S.; Bello, J.P. BirdVoxDetect: Large-Scale Detection and Classification of Flight Calls for Bird Migration Monitoring. IEEE/ACM Trans. Audio Speech Lang. Process. 2024, 32, 4134–4145. https://doi.org/10.1109/TASLP.2024.3444486.
100. Michez, A.; Broset, S.; Lejeune, P. Ears in the sky: Potential of drones for the bioacoustic monitoring of birds and bats. Drones 2021, 5, 9. https://doi.org/10.3390/drones5010009.
101. Guerrero, M.J.; Bedoya, C.L.; López, J.D.; Daza, J.M.; Isaza, C. Acoustic animal identification using unsupervised learning. Methods Ecol. Evol. 2023, 14, 1500–1514. https://doi.org/10.1111/2041-210X.14103.
102. Neethirajan, S. ChickTrack—A quantitative tracking tool for measuring chicken activity. Measurement 2022, 191, 110819. https://doi.org/10.1016/j.measurement.2022.110819.
103. Berthet, M.; Coye, C.; Dezecache, G.; Kuhn, J. Animal linguistics: a primer. Biol. Rev. 2023, 98, 81–98. https://doi.org/10.1111/brv.12897.
104. Hagiwara, M.; Hoffman, B.; Liu, J.-Y.; Cusimano, M.; Effenberger, F.; Zacarian, K. BEANS: The Benchmark of Animal Sounds. arXiv 2022, arXiv:2210.12300. https://doi.org/10.48550/arXiv.2210.12300.
105. Banbury, C.R.; Reddi, V.J.; Lam, M.; Fu, W.; Fazel, A.; Holleman, J.; Huang, X.; Hurtado, R.; Kanter, D.; Lokhmotov, A.; et al. Benchmarking TinyML Systems: Challenges and Direction. arXiv 2020, arXiv:2003.04821. https://doi.org/10.48550/arXiv.2003.04821.

106. 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, 147–168. https://doi.org/10.15388/22-INFOR505.
107. Goyal, V.; Yadav, A.; Mukherjee, R. A Literature Review on the Role of Internet of Things, Computer Vision, and Sound Analysis in a Smart Poultry Farm. ACS Agric. Sci. Technol. 2024, 4, 368–388. https://doi.org/10.1021/acsagscitech.3c00467.
108. Ginovart-Panisello, G.J.; Alsina-Pagès, R.M.; Iriondo Sanz, I.; Panisello Monjo, T.; Call Prat, M. 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 2020, 20, 4732. https://doi.org/10.3390/s20174732.
109. Ginovart-Panisello, G.-J.; Alsina-Pagès, R.M.; Panisello Monjo, T. Acoustic description of bird broiler vocalisations in a real-life intensive farm and its impact on animal welfare: A comparative analysis of recordings. Eng. Proc. 2020, 2, 53. https://doi.org/10.3390/ecsa-7-08164.
110. Ginovart-Panisello, G.J.; Iriondo Sanz, I.; Panisello Monjo, T.; Riva, S.; Garriga Dicuzzo, T.; Abancens Escuer, E.; Alsina-Pagès, R.M. Trend and representativeness of acoustic features of broiler chicken vocalisations related to CO2. Applied Sciences 2022, 12, 10480. https://doi.org/10.3390/app122010480.
111. Niu, X.; Jiang, Z.; Peng, Y.; Huang, S.; Wang, Z.; Shi, L. Visual cognition of avians and its underlying neural mechanism: a review. Avian Res. 2022, 13, 100023. https://doi.org/10.1016/j.avrs.2022.100023.
112. Mutanu, L.; Gohil, J.; Gupta, K.; Wagio, P.; Kotonya, G. A review of automated bioacoustics and general acoustics classification research. Sensors 2022, 22, 8361. https://doi.org/10.3390/s22218361.
113. Coutant, M.; Villain, A.S.; Briefer, E.F. A scoping review of the use of bioacoustics to assess various components of farm animal welfare. Appl. Anim. Behav. Sci. 2024, 275, 106286. https://doi.org/10.1016/j.applanim.2024.106286.
114. Van Merriënboer, B.; Hamer, J.; Dumoulin, V.; Triantafillou, E.; Denton, T. Birds, bats and beyond: evaluating generalization in bioacoustics models. Front. Bird Sci. 2024, 3, 1369756. https://doi.org/10.3389/fbirs.2024.1369756.
115. Stowell, D. Computational bioacoustics with deep learning: a review and roadmap. PeerJ 2022, 10, e13152. https://doi.org/10.7717/peerj.13152.
116. Bolhuis, J.J.; Beckers, G.J.L.; Huybregts, M.A.C.; Berwick, R.C.; Everaert, M.B.H. Meaningful syntactic structure in songbird vocalizations? PLOS Biol. 2018, 16, e2005157. https://doi.org/10.1371/journal.pbio.2005157.
117. Jarvis, E.D. Evolution of vocal learning and spoken language. Science 2019, 366, 50–54. https://doi.org/10.1126/science.aax0287.
118. Takeshita, M.; Rzepka, R. Speciesism in natural language processing research. AI Ethics 2024, 5, 2961–2976. https://doi.org/10.1007/s43681-024-00606-3.

119. Marino, L. Thinking chickens: a review of cognition, emotion, and behavior in the domestic chicken. Anim. Cogn. 2017, 20, 127–147. https://doi.org/10.1007/s10071-016-1064-4.
120. Loizou, P.C. Speech Enhancement: Theory and Practice, 2nd ed.; CRC Press: Boca Raton, FL, USA, 2013. https://doi.org/10.1201/b14529.
121. European Parliament and Council. Regulation (EU) 2016/679 (General Data Protection Regulation); Official Journal of the European Union, L119. Available online: https://eur-lex.europa.eu/eli/reg/2016/679/oj (27 June 2025).
122. Gemtou, M.; Casares Guillén, B.; Anastasiou, E. Smart farming technologies and sustainability. In Digital Sustainability; Lynn, T., Rosati, P., Kreps, D., Conboy, K., Eds.; Palgrave Macmillan: Cham, Switzerland, 2024; pp. 99–120. https://doi.org/10.1007/978-3-031-61749-2_6.
123. Snell, J.; Swersky, K.; Zemel, R.S. Prototypical networks for few-shot learning. In Advances in Neural Information Processing Systems; Curran Associates, Inc.: Red Hook, NY, USA, 2017; Vol. 30 . Available online: https://github.com/jakesnell/prototypical-networks (27 June 2025).
124. Selvaraju, R.R.; Cogswell, M.; Das, A.; Vedantam, R.; Parikh, D.; Batra, D. Grad-CAM: Visual explanations from deep networks via gradient-based localization. In Proceedings of the 2017 IEEE International Conference on Computer Vision, Venice, Italy, 22–29 October 2017; pp. 618–626. https://doi.org/10.1109/ICCV.2017.74.
125. Coffey, K.R.; Marx, R.G.; Neumaier, J.F. DeepSqueak: A Deep Learning-Based System for Detection and Analysis of Ultrasonic Vocalizations. Neuropsychopharmacology 2019, 44, 859–868. https://doi.org/10.1038/s41386-018-0303-6.

Appendix

This appendix compiles the technical resources, methodologies, and supplementary information that underpin the experimental and computational work presented in this thesis. It serves as a detailed reference point for libraries, functions, model architectures, statistical procedures, publication permissions, and the author's contributions to scholarly literature. Such documentation ensures transparency, reproducibility, and facilitates future extensions of this research.

Appendix A

Table A.1: Core Python Libraries and Their Purposes in Data Processing and Machine Learning

Library	Purpose
pandas	Data manipulation, loading CSVs
numpy	Numeric operations, array manipulation
matplotlib.pyplot	Plotting (confusion matrices, metrics)
seaborn	Enhanced visualization (heatmaps, barplots)
scikit-learn (sklearn)	ML pipeline: preprocessing, models, metrics, splits
catboost	Gradient boosting ensemble classifier
pytorch_tabnet	TabNet model for tabular data classification
statsmodels	McNemar's test for contingency table comparison
scipy.stats	Chi-square tests
pickle	Saving/loading fitted models

Appendix B

Table B.1: Machine Learning Classifiers and Preprocessing Tools with Their Source Libraries

Classifier / Model	**Library / Source**
Randomforestclassifier	*sklearn.ensemble*
Gradientboostingclassifier	*sklearn.ensemble*
Adaboostclassifier	*sklearn.ensemble*
Extratreesclassifier	*sklearn.ensemble*
Catboostclassifier	*catboost*
Histgradientboostingclassifier	*sklearn.ensemble*
Mlpclassifier	*sklearn.neural_network*
Tabnetclassifier	*pytorch_tabnet*
Standardscaler	*sklearn.preprocessing*
Train_Test_Split	*sklearn.model_selection*

Appendix C

Table C.1: Functions and Methods for Evaluation, Statistical Testing, and Pipeline Operations

Function / Method	**Purpose / Notes**
confusion_matrix, accuracy_score, precision_score, recall_score, f1_score, log_loss, cohen_kappa_score	Compute standard performance metrics
train_test_split	Stratified splitting of dataset into train/test
standardscaler.fit_transform	Feature scaling to zero mean, unit variance
plot_confusion_matrix()	Custom function to visualize confusion matrices
plot_metrics()	Plot bar charts for accuracy, precision, recall, F1 score
compare_models()	Compute McNemar, Kappa, weighted F1, log loss diff, chi-square
mcnemar (statsmodels)	Statistical significance test for paired nominal data
chi2_contingency (scipy)	Chi-square test on confusion matrices
pickle.dump / load	Save/load fitted ensemble of classifiers for reuse

Appendix D

Table D.1: Libraries, Classes, and Functions Used in Audio Preprocessing, Parallel Execution, and Transformer-Based ASR

Category	**Details**
python libraries	*os, librosa, pydub, concurrent.futures (ThreadPoolExecutor), uuid, torch, transformers*
classes	*AudioSegment (pydub), Wav2Vec2Processor, Wav2Vec2ForCTC*
functions / methods	*AudioSegment.from_file, set_frame_rate, set_channels, exportlibrosa.load, processor and model inference via Wav2Vec2Processor / Wav2Vec2ForCTCThreadPoolExecutor.submit, as_completed for concurrencytorch.argmax for decoding logits, softmax for probabilities*
parallelization	*Batch-processing of audio files with ThreadPoolExecutor, accelerating I/O and waveform segmentation*
device utilization	*CUDA-enabled inference for Wav2Vec2 ASR*
file management	*Temporary segment files auto-removed using os.remove, unique IDs via uuid.uuid4*

Appendix E

Table E.1: Libraries and Functions Used in NLP-Based Sentiment, Phonetic, and Frequency Analysis

Category	**Details**
Python libraries	*nltk, transformers, torch, collections.Counter, matplotlib.pyplot*
Classes	*BertTokenizer, BertForSequenceClassification*
Functions / Methods	*nltk.word_tokenize for unigramsCounter(...) for frequency distributionsBertTokenizer and BERT inference for sentiment logitstorch softmax + classification*
Special Analysis	*Pie charts for sentiment proportions, bar plots for vowel and character frequency, bigram analysis using nltk.ngrams*
Visualization	*matplotlib.pyplot, seaborn for distribution plots*
Device Utilization	*Sentiment predictions performed on CUDA GPUs with BERT*

Appendix F

Table F.1: Libraries, Classes, and Functions for BERT-Based Supervised Classification and Statistical Comparisons

Category	**Details**
Python libraries	*pandas, numpy, torch, sklearn (train_test_split, StandardScaler, classification_report, confusion_matrix), matplotlib, seaborn*
Classes	*BertModel, BertConfig, AdamW, DataLoader, Dataset, CrossEntropyLoss, get_linear_schedule_with_warmup*
Functions / Methods	*forward on BertModel using inputs_embeds, optimizer.step, scheduler.step, evaluation with classification_report, confusion_matrix, seaborn heatmaps*
Training Workflow	*Full multi-epoch supervised training loop on CUDA GPUs for BERT classification, with validation loss, accuracy, and confusion matrix visualizations*
Special Statistical Tests	*McNemar, Cohen's Kappa, Chi-Square, Matthews Correlation Coefficient, log_loss comparisons across models (for traditional ML classifiers on MFCC features)*

Appendix G: Permission to use

In presenting this thesis in partial fulfilment of the requirements for the master's in computer science degree from Dalhousie University, I agree that the Libraries of this University may make it freely available for inspection. I further agree that permission for copying of this thesis in any manner, in whole or in part, for scholarly purposes may be granted by the professor or professors who supervised my thesis work or, in their absence, by the Head of the Department or the Dean of the College in which my thesis work was done. It is understood that any copying or publication or use of this thesis or parts thereof for financial gain shall not be allowed without my written permission. It is also understood that due recognition shall be given to me and to Dalhousie University in any scholarly use which may be made of any material in my thesis.

Requests for permission to copy or to make other use of the material in this thesis

in whole or part should be addressed to:

Head of the Faculty of Computer Science 6050 University Ave,

Dalhousie University,

Halifax, Nova Scotia, Canada B3H 1W5

Appendix H: My Publications

1. V. Manikandan and S. Neethirajan, "AI-powered vocalization analysis in poultry: Systematic review of health, behavior, and welfare monitoring," Sensors, vol. 25, no. 13, Art. no. 4058, Jun. 2025. [Online]. Available: https://doi.org/10.3390/s25134058
Note: This article was also featured as part of the Special Issue *"Feature Papers in Smart Agriculture 2025"* in *Sensors*, highlighting advanced research in smart farming and precision livestock monitoring.

2. V. Manikandan and S. Neethirajan, "Decoding poultry welfare from sound—A machine learning framework for non-invasive acoustic monitoring," Sensors, vol. 25, no. 9, Art. no. 2912, May 2025. [Online]. Available: https://doi.org/10.3390/s25092912

3. V. Manikandan and S. Neethirajan, "Decoding poultry vocalizations: Natural language processing and transformer models for semantic and emotional analysis," bioRxiv, preprint, Dec. 2024. [Online]. Available: https://doi.org/10.1101/2024.12.18.629057. Submitted and under Review.