Journal of Agricultural Engineering

https://www.agroengineering.org/

Acoustic feature analysis of normal and abnormal calls of white-feathered broilers

Bowen Sun,[1,2*] Xiangchao Kong,[1,2*] Xingkai Peng,[1,2] Changxi Chen,[1,2] Wanchao Zhang,[1,2] Guangyu Zhao,[1,2] Yongmin Guo,[1,2] Kaisi Yang[1,2]

[1]College of Computer and Information Engineering, Tianjin Agricultural University, Tianjin

[2]Key Laboratory of Smart Farming, Ministry of Agriculture and Rural Affairs, Tianjin, China

[*] *These authors share first-authorship.*

Corresponding author: Changxi Chen, College of Computer and Information Engineering, Tianjin Agricultural University, Tianjin 300384, China. E-mail: chenchangxi@tjau.edu.cn

Please cite this article as doi: 10.4081/jae.2025.1684

Submitted: 1 May 2025
Accepted: 3 November 2025

Acoustic feature analysis of normal and abnormal calls of white-feathered broilers

Bowen Sun,[1,2*] Xiangchao Kong,[1,2*] Xingkai Peng,[1,2] Changxi Chen,[1,2] Wanchao Zhang,[1,2] Guangyu Zhao,[1,2] Yongmin Guo,[1,2] Kaisi Yang[1,2]

[1]College of Computer and Information Engineering, Tianjin Agricultural University, Tianjin

[2]Key Laboratory of Smart Farming, Ministry of Agriculture and Rural Affairs, Tianjin, China

[*] *These authors share first-authorship.*

Corresponding author: Changxi Chen, College of Computer and Information Engineering, Tianjin Agricultural University, Tianjin 300384, China. E-mail: chenchangxi@tjau.edu.cn

Contributions: all authors made a substantive intellectual contribution, read and approved the final version of the manuscript and agreed to be accountable for all aspects of the work.

Conflict of interest: the authors declare no competing interests, and all authors confirm accuracy.

Abstract

This study investigates the acoustic characteristics of normal and abnormal calls in white-feathered broilers to propose a method for early detection of non-healthy conditions. Vocalizations were collected from 2-week-old broilers over a 21-day period and analyzed using time-domain and frequency-domain features, including maximum amplitude, effective amplitude, fundamental frequency, and pulse index. Significant differences were identified between normal calls and abnormal calls influenced by laryngeal mucus, with support vector machines and random forest classifiers achieving accuracies of 97.8% and 98.76%, respectively. Unlike previous empirical feature aggregation methods, this research employs statistically validated feature selection aligned with physiological mechanisms, enhancing interpretability and performance. The proposed framework offers a practical, automated solution for on-farm monitoring of broiler vocalizations, contributing to early detection of abnormal signs and improved management in precision poultry farming.

Key words: Acoustic features; health monitoring; model; sound monitoring; white-feathered broilers.

Introduction

Animal vocalizations contain rich information that can effectively reflect their health and production status, which is crucial for building an environment suitable for animal welfare (Wei *et al*., 2023). Traditionally, the health assessment of white-feathered broilers has relied on visual observation and physiological indicators, which suffer from subjectivity and delayed warnings (Yao *et al*., 2023). More effective detection methods are urgently needed, especially to address the issue of limited acoustic feature parameters currently used in research, aiming to statistically identify key acoustic features that can significantly distinguish different health states of broilers.

Although significant progress has been made in sound detection for broiler health monitoring, with broad application prospects, direct studies on the vocalizations of white-feathered broilers are relatively rare (Bhardwaj *et al*., 2022). There is also a lack of systematic comparative analysis of their sound features, which has led to empirical feature selection. Tao *et al* .(2022) extracted 60 acoustic features from time-domain, frequency-domain, MFCC, and sparse representations, and selected 30 features with the highest contribution to classification performance using the random forest algorithm. This method tested multiple segments of broiler vocal signals and achieved a recognition accuracy of 99.12%, demonstrating the potential of the model for automated broiler vocal recognition. Alireza *et al*. (2021) studied the application of five acoustic features in bird vocalizations for health monitoring, finding that wavelet entropy was most effective in detecting infectious bronchitis. For Newcastle disease, wavelet entropy and Mel-frequency cepstral coefficients (MFCC) had similar detection accuracies, 80% and 78%, respectively, but wavelet entropy was more reliable in detecting healthy birds. Cuan *et al*. (2020) proposed a method based on convolutional neural networks for chicken sound recognition to detect avian influenza infection. They extracted four types of acoustic features: logarithmic bank features, MFCC, and their first and second-order differences, and combined them with a CNN model for classification. The model achieved accuracy rates of 93.01%, 95.05%, and 97.43% on days 2, 4, and 6 after virus injection, respectively, proving its rapid and effective detection capabilities. Du *et al*. (2021) developed a method based on a three-stimulus-resonance peak model for classifying chicken vocalizations. They extracted three stimuli and MFCC features and used BP neural networks and Gaussian mixture models for classification, achieving an average accuracy of 94.9% with a 12-dimensional MFCC and BPNN feature model. Li *et al*. (2022)

designed a gender detection model for chicks based on deep learning, distinguishing male and female chicks by vocal differences at one day of age, achieving a detection accuracy of 95% with the ResNet-50 model.

Many studies have focused on traditional features such as MFCC and linear predictive cepstral coefficients (Abdul *et al*., 2022), which are widely used in acoustic analysis. While these features are classical and widely applied, they often fail to fully capture the complex health information in vocal signals, especially subtle but crucial changes. Further exploration and development of more comprehensive feature parameters are necessary to improve the accuracy and reliability of disease detection.

In fact, broiler vocalizations contain abundant acoustic features (Curtin *et al*., 2014), including, but not limited to, spectral features, time-envelope features, nonlinear features (such as bispectral, higher-order cumulants), and time-frequency combined features (Mahdavian *et al*., 2021). These features have distinct characteristics in describing different dimensions of sound signals, some sensitive to frequency changes, and others better reflecting temporal or nonlinear aspects (Xu *et al*., 2020). Therefore, relying solely on a few or single acoustic features for model training limits the recognition ability and generalization performance of the model, making it difficult to adapt to the diversity of vocalizations across different breeds, ages, and health conditions of broilers. Given the shortcomings of current sound detection in broiler health monitoring, it is necessary to explore more diversified feature extraction approaches and continue to improve and optimize the related technical systems (Huang *et al*., 2021).

To sum up, the current body of research on livestock and poultry acoustics has laid a critical foundation for the use of vocalizations in health monitoring. Studies focusing on classic features such as mel-frequency cepstral coefficients (MFCCs) and linear predictive cepstral coefficients (LPCCs) have undeniably demonstrated the feasibility and value of audio-based bioacoustic analysis. These approaches are computationally efficient, well-understood, and have provided significant benchmarks in distinguishing broad health states, thereby establishing a strong basis for non-invasive monitoring. Furthermore, the work of researchers like Tao and Cuan has been instrumental in proving the high classification accuracy achievable by combining multiple feature sets with powerful machine learning models, moving the field firmly beyond purely observational methods.

However, despite these valuable contributions, certain limitations persist. The heavy reliance on these traditional feature sets, while practical, often fails to fully capture the complexity and subtlety of health-related changes in vocalizations. Many studies employ a relatively narrow range of feature parameters, which can constrain the model's ability to generalize across diverse conditions, such as different breeds or specific illnesses that manifest through nuanced acoustic shifts. These challenges not only impact the accuracy and robustness of broiler health monitoring but may also limit the broader application and scalability of sound detection technologies in precision livestock farming. Therefore, while building upon the solid groundwork established by previous research, there is a compelling need to explore more comprehensive and discriminatory feature extraction strategies to overcome these limitations and enhance the effectiveness of vocalization-based health assessment.

This study selected 60 white-feathered broiler chickens at 2 weeks of age and continuously recorded their vocalizations for 21 days, with a nearly equal gender ratio of approximately 1:1. Using microphones, their normal and abnormal calls were recorded at a fixed time each afternoon throughout this period, with each recording lasting for 10 minutes. The abnormal calls were classified as whining sounds and cackling sounds through manual identification. Quantitative analysis was conducted on these three types of sounds, and acoustic features such as duration, amplitude, and frequency were extracted and analyzed using independent sample T-tests. This study aims to explore and determine unique acoustic features that effectively distinguish between healthy and abnormal states, and validate their performance in machine learning models, in order to develop a model that can effectively distinguish abnormal calls from healthy calls, providing new quantitative basis for early health warning of broiler chickens. The acoustic-based health monitoring method overcomes the subjectivity and delay inherent in traditional assessments, provides valuable features for machine learning models, and demonstrates significant application potential.

Materials and Methods

Sound signal collection

This study focuses on WOD168 white-feathered broiler chickens. The experiment was conducted at the Simate Broiler Farm (GPS coordinates: 37.17°N, 116.44°E) in Pingyuan County, Shandong

Province, China. The poultry house used in this study was a mechanically ventilated building with a length of 101 m, width of 16 m, eave height of 4.2 m, and ridge height of 6 m. Its cross-sectional area was 81.6 m^2, yielding an internal volume of 8,241.6 m^3. The housing system consisted of 8 cage rows, each comprising 73 cages per row and 4 tiers per row, with 30 broilers per cage, resulting in a total stocking capacity of 70,080 birds (8 rows × 73 cages × 4 tiers × 30 birds). This structural and stocking information provides context for the experimental scale and ensures reproducibility.

A total of 60 broilers were selected at 2 weeks of age from the same batch, comprising 40 healthy and 20 unhealthy birds, with a sex ratio close to 1:1. The birds were housed in a dedicated experimental area within the main poultry house, which was physically isolated from the commercial flock by solid partitions to maintain standard environmental conditions (e.g., ventilation, temperature, humidity, and lighting). To adhere to the standard commercial stocking density of a maximum of 30 birds per cage and to ensure separation by health status, the birds were allocated into three adjacent cages. Specifically, the 40 healthy birds were divided into two groups of 20 birds each and housed in two separate cages, while all 20 unhealthy birds were housed together in a third cage. All cages were of the same standard type used throughout the facility. Feed and water were provided ad libitum via the central automated system. This configuration ensured that the collected acoustic data reflected natural vocalization behaviors under representative farming conditions while enabling clear distinction between health statuses and minimizing disturbance from routine operations.

Vocalizations were recorded from the three cages sequentially over 21 consecutive days using a XinKe V-01 microphone with a magnetic base. Each day, the microphone was deployed at a fixed, identical position relative to each cage in turn. Specifically, for each recording session, the microphone was attached to the outer side of the cage rack at the third tier, centered on the target cage, with the microphone head positioned at a height of approximately 1.5 meters above the floor. This standardized placement ensured consistent acoustic recording conditions across all cages, prevented physical contact with the birds, and minimized behavioral disturbance. The microphone has a sensitivity of -36±3 dB, a signal-to-noise ratio ≥70 dB, and a frequency response range of 50 Hz-12 kHz. All recordings were made with a 16-bit depth and a sampling rate of 16 kHz. Recording sessions for each cage were conducted daily at a fixed time in the

afternoon, with each session lasting 10 min. To avoid capturing stress-induced vocalizations and to ensure the sounds reflected natural conditions, the microphone was mounted externally without direct exposure to the chickens during all recordings.

The collected sounds were classified by two experienced professional breeders through manual auditory screening based on acoustic morphology and context. The classification followed established ethological criteria adapted from prior poultry vocalization studies (de Carvalho Soster *et al*., 2025). Normal vocalizations were clear. Abnormal vocalizations included both rhythmic clucking calls and mucus-associated whimpering calls, which were characterized by wheezing, gurgling, or strained tones and often occurred in series with irregular rhythm. Any ambiguous cases were re-evaluated jointly until consensus was reached.

Sound signal preprocessing

The recorded chicken vocalizations contain significant amounts of irrelevant information and noise (Catania *et al*., 2013), which must be removed through preprocessing in order to extract meaningful sound features (Biocca *et al*., 2019). The main preprocessing steps in this study include pre-emphasis, denoising, and frame windowing (Akçay *et al*., 2020).

Pre-emphasis is a high-pass filter that processes the signal using a first-order filter to emphasize high-frequency components while reducing low-frequency components (Dong *et al*., 2020). The process involves subtracting the previous sample multiplied by a coefficient α (set to 0.95 in this study) from the current sample.

Since sound signals are highly complex and non-stationary, it is generally assumed that the characteristics of sound signals remain stable or change slowly within short time segments. Since the audio signals are usually complex and non-stationary, it is generally believed that their characteristics remain relatively stable over a longer period of time, or change at a relatively slow rate. In this study, when choosing the frame length, we referred to the observation results, which indicated that the calls of broiler chickens typically change within a time scale of 15 to 60 milliseconds in a short period of time. Therefore, a frame length of 50 milliseconds (equivalent to 800 samples at a 16 kHz sampling rate) was selected. As shown in Figure 1, a frame length of 50 milliseconds can achieve a balance between temporal resolution and feature representation, enabling a more comprehensive capture of the acoustic features of short-term sound events. To

ensure a smooth transition between frames, a 50% (25 milliseconds) overlap method was adopted. The effective signal length of each recording was 5 seconds, containing multiple analysis frames. In order to reduce edge discontinuities and spectral leakage, a Hamming window is used in this study. The equation for the Hamming window is as follows:

$$w(n) = 0.54 - 0.46\cos(N - 12\pi n) \qquad \text{(Eq. 1)}$$

where n is the sample index within the window ($0 \leq n \leq N\text{-}1$), and N is the length of the window (i.e., the number of samples, N=800 in our implementation). The coefficients 0.54 and 0.46 in the Hamming window are used to adjust the relative size of the main lobe and side lobes, providing approximately -43 dB sidelobe suppression.

The short-time Fourier transform (STFT) is then applied to convert the windowed audio signal into a time-frequency domain spectrum, extracting its amplitude, phase information, and power spectrum. A 1024-point FFT was used to compute the STFT, providing a frequency resolution of 15.625 Hz. To estimate the noise energy, the average of the first 30 frames (approximately 1.5 seconds) of the audio signal, which contained only background noise, was calculated.

To reduce spectral discontinuities, the amplitude spectrum is smoothed using a moving average filter with a 3-band width. By combining the estimated noise energy with the signal's power spectrum, spectral subtraction is employed for denoising (Balaji *et al*., 2020). This step adjusts the amount of noise energy subtracted (using an over-subtraction factor of 3.0) to balance the denoising effect and signal fidelity while controlling small values in the amplitude spectrum to enhance signal smoothness. The enhanced amplitude spectrum is then combined with the original phase information, and the inverse STFT is applied to restore the time-domain audio signal.

Sound signal endpoint detection

For chicken vocalization detection, this study proposes an adaptive dual-threshold endpoint detection algorithm based on short-time energy (E) and short-time zero-crossing rate (Z). The threshold coefficients (20% and 30%) were empirically optimized through iterative testing on a subset of 50 recordings to maximize the F1-score for vocalization detection, a common approach in bioacoustic studies (Yin *et al*., 2024). To constrain the vocal pulse duration, the minimum length was set to 6 frames (corresponding to 300 ms at our 50 ms frame length), while the

maximum length was set to 30 frames (corresponding to 1500 ms or 1.5 s). These limits effectively filtered out short-duration impulsive noises and long-duration non-vocal events such as equipment hum.

The algorithm proceeds as follows: First, the moving averages of E and Z (MA_E and MA_Z) are computed over a 500 ms window to adapt to varying background conditions. A candidate start point is identified when $E > 0.2 * MA\_E$ and $Z > 0.2 * MA\_Z$ simultaneously. This candidate is confirmed as a true start point if, within the subsequent 200 ms, both $E > 0.3 * MA\_E$ and $Z > 0.3 * MA\_Z$ are satisfied. After start point confirmation, the vocal endpoint is identified when both E and Z fall below $0.3 * MA\_E$ and $0.3 * MA\_Z$ for more than 6 consecutive frames.

Figure 2 illustrates the endpoint detection process on a representative audio clip containing both normal and abnormal calls, demonstrating the effectiveness of the chosen thresholds.

Sound signal feature extraction

Following speech denoising and endpoint detection on signals containing three types of vocalizations, this study systematically extracts features in both time and frequency domains (Lin *et al.*, 2001). The extracted features and their mathematical formulations are summarized in Table 1.

Time-domain feature extraction

Time-domain features are calculated directly from the signal's amplitude waveform, reflecting its energy distribution and statistical properties. The standard extraction process involves segmenting the signal into short frames, applying a Hamming window to each frame to minimize spectral leakage, and calculating features within each window. Average values are then computed across all frames to obtain the final feature vector.

A suite of time-domain features is extracted to capture the signal's instantaneous characteristics and overall energy profile. These include maximum amplitude, peak value, root mean square (RMS) amplitude, variance, skewness, Kurtosis, peak-to-peak value, crest factor, margin index, pulse index, and waveform index. Their mathematical formulations and descriptions are detailed in Table 1. These features are instrumental in detecting impulsive components, quantifying the dynamic range, and assessing the waveform's similarity to a pure sinusoid.

Frequency-domain features are derived by transforming the windowed time-domain signal into the frequency domain using a Fourier Transform. The transformation for a discrete-time signal $x[n]$ of length N is given by:

$$X(f) = \sum_{n=1}^{N-1} x(n) \cdot e^{-j2\pi \frac{fn}{N}} \quad \text{(Eq. 2)}$$

Where:

$X(f)$ is the frequency-domain signal

$x(n)$ is the time-domain signal

N is the signal length,

f is the frequency.

The extracted frequency-domain features provide insights into the spectral structure and variability of the sound signal. These include the Fundamental Frequency (F0), which corresponds to the perceived pitch; Spectral Centroid, indicating the brightness of the sound; Spectral Bandwidth and Spectral Roll-off, describing the spread of spectral energy; and RMS Frequency. Their precise mathematical definitions are provided in Table 1. These features are crucial for analyzing the harmonic content, resonant frequencies, and spectral shape of the vocalization.

Extraction of traditional acoustic features

Figure 3 illustrates the extraction process of mel-frequency cepstral coefficients (MFCC) and linear prediction cepstral coefficients**.** MFCC and LPCC are two fundamental cepstral feature extraction techniques. The overall workflow for both methods is summarized in the flowchart below, highlighting their parallel yet distinct processes.

MFCC extraction is designed to model the human auditory system. The signal undergoes pre-emphasis and is split into short, windowed frames. The Fourier transform converts these frames into the frequency domain. The resulting spectrum is then processed through a Mel-scaled filter bank to simulate nonlinear human hearing, followed by a logarithm to compress dynamic range. Finally, the discrete cosine transform (DCT) decorrelates the filter bank outputs to produce the final cepstral coefficients, with the first 13 typically retained as features. In contrast, LPCC extraction focuses on modeling the physical properties of the vocal tract. After pre-emphasis, framing, and windowing, it analyzes the signal's autocorrelation to perform linear predictive

coding (LPC). The LPC coefficients, which model the vocal tract filter, are then converted into cepstral coefficients via a recursive formula, with the first 13 coefficients retained as the LPCC features. The distinct principles and applications of these two feature sets are concisely compared in the Table 2.

Statistical analysis and testing

The experiment was conducted on a Windows 10 64-bit operating system using Python 3.10 for audio processing (utilizing libraries such as librosa, scipy, and numpy) and SPSS software for statistical analysis. The core innovation of our feature selection strategy lies in the rigorous statistical screening of a comprehensive set of acoustic features, rather than relying on empirical or conventional subsets. To this end, an independent-samples t-test was conducted to analyze the differences in acoustic feature parameters between normal calls, whimpering calls, and clucking calls (Yu *et al*., 2022). This test served as a critical filter to identify and select features with significant discriminatory power for subsequent model training, directly addressing the feature dimensionality and selection challenge. A *p*-value of less than 0.05 was set as the threshold for statistical significance.

Following this statistically-guided feature selection, multiple machine learning models were employed for vocalization classification, including support vector machines (SVM), random forests, naive bayes, and K-nearest neighbors (KNN). Each model offers distinct advantages: SVM constructs an optimal hyperplane for high accuracy and minimizes overfitting; random forest integrates multiple decision trees for robustness against complex nonlinear relationships; naive bayes, based on conditional independence assumption, is efficient for large datasets; and KNN offers intuitive classification based on nearest neighbors (Balaji *et al*., 2021). The use of these diverse models allows for a robust evaluation of the effectiveness of the selected feature subset.

Results

Feature selection based on independent-samples *t*-test

The independent-samples *t*-test provided a data-driven foundation for feature selection. As detailed in Table 3, the analysis revealed statistically significant differences ($p<0.05$) in all sixteen examined acoustic features when comparing normal calls to both whimpering and clucking calls.

This result indicated that every feature in our extracted set possessed significant discriminatory power for this specific classification task. Therefore, based on this rigorous statistical evaluation, the entire set of features was selected for subsequent machine learning model training. The profoundly low *p*-values (often $p<0.001$) and substantial t-statistics across time-domain (e.g., maximum amplitude, variance, pulse index), frequency-domain (e.g., fundamental frequency, spectral line width), and spectral features (resonance peaks) demonstrate that these parameters collectively capture essential and significant aspects of vocal differences, likely reflecting underlying physiological or behavioral states. This comprehensive, statistically-vetted feature set forms the basis of our model input, ensuring that only features with proven discriminative power are utilized.

Model performance evaluation

The performance of the machine learning models, trained on the statistically-selected feature set, is summarized in Table 4. The random forest algorithm demonstrated exceptional classification capability, achieving near-perfect scores across all evaluation metrics (accuracy, precision, recall, F1 score$\approx$0.988). This outstanding result highlights the synergistic effect of combining a powerful classifier with a feature set pre-validated for high discriminative power. Furthermore, models utilizing the t-test selected features consistently and significantly outperformed those employing conventional cepstral features (MFCCs, LPCCs) alone or in combination. This performance gap underscores the value of our comprehensive feature selection methodology over traditional approaches that rely on a predetermined, limited set of features.

Among the cepstral features, LPCCs generally showed superior predictive capability compared to MFCCs. Combining MFCC and LPCC features led to a performance improvement over using either set independently, suggesting complementary information. However, even this combination failed to match the performance achieved by the broader, statistically-selected feature set, emphasizing the necessity of incorporating a diverse range of acoustic characteristics for optimal classification performance in this domain.

Discussion

The accumulation of mucus in the larynx may cause changes in the vibration characteristics of

the vocal cords, thus affecting the time domain and frequency domain characteristics of the sound. Specifically, the significant difference between normal and whimper calls suggests that the presence of mucus may have reduced the amplitude of vocal cord vibrations, which in turn affected the intensity and clarity of the sound. From a time-domain perspective, the differences between normal calls and whimpering calls are particularly evident in parameters such as maximum amplitude and effective amplitude. For instance, the maximum amplitude of normal calls is typically lower (around 0.685±0.113), whereas the maximum amplitude of whimpering calls is significantly reduced (around 0.503±0.094). This change suggests that throat mucus may reduce the amplitude of vocal cord vibrations, leading to an overall loss of sound energy, which in turn affects sound quality and clarity (Peters *et al*., 2021). The significant decrease in effective amplitude (from 0.108 to 0.030) further supports this idea, reflecting the suppressive effect of mucus on vocal cord vibrations. Figure 4 highlights the waveform variations for different call types. In terms of frequency-domain features, the significant differences in fundamental frequency and resonance peaks further reveal the changes in sound quality. The accumulation of mucus in the throat likely alters the mass and tension of the vocal cords, resulting in a decrease and instability in the fundamental frequency (Hegde *et al*., 2024). For example, the fundamental frequency of normal calls is 501.395±319.346 Hz, while the fundamental frequency of whimpering calls is significantly lower (145.674±0.280 Hz).This frequency change may be due to altered vocal cord vibration frequencies, manifesting as instability and lower frequencies (Li *et al*., 2021). Spectrograms (Figure 5) visually represent the energy distribution across different frequencies for these call types, where mucus-influenced calls show broader and less distinct frequency bands. These findings are consistent with the research by Moura, who noted that the frequency and amplitude of bird calls can reflect their thermal comfort, emphasizing the impact of physiological health on vocalizations (Moura *et al*., 2008). Additionally, the acoustic features between normal calls and clucking calls also demonstrate changes in sound quality, particularly in terms of fundamental frequency and resonance peaks. Clucking calls exhibit a significantly higher maximum amplitude than normal calls (from 0.685 to 0.802), but their fundamental frequency decreases markedly (from 501.395 Hz to 146.844 Hz). This phenomenon suggests a connection between whimpering calls and clucking calls, such as pneumonia, which typically lead to instability in sound energy and significant frequency changes (Yadav *et al*., 2020). In terms

of bandwidth, which reflects the spectral spread of sound, calls with mucus accumulation show a significant increase in bandwidth, suggesting greater spectral line width. This is further corroborated by the bandwidth bar chart (Figure 6), which highlights this increase across the different call types.

Compared to other studies, this research also achieved positive results in feature selection and model evaluation. For instance, research by Bu, which analyzed the sounds of seven woodboring pests, proposed a pulse-based acoustic feature extraction method, which differs from the traditional time-domain and frequency-domain methods employed in this study (Bu *et al*., 2016). By focusing on discrete pulse features, the pest monitoring system proposed by Bu was able to identify pests at an early stage, offering a unique perspective on animal sound data analysis. Similarly, research by Sun highlighted subtle changes in the pulse groups of cricket calls, emphasizing the importance of minute differences between pulses in species identification (Sun *et al*., 2019). These studies have enriched the methodology for analyzing animal acoustic features and provided valuable insights for this research.

Overall, the significant differences in maximum amplitude, effective amplitude, fundamental frequency, and pulse index between normal calls, whimpering calls and clucking calls reflect changes in sound intensity and quality. Notably, throat mucus reduces the amplitude of vocal cord vibrations and leads to a decrease in both sound energy and clarity. Similarly, the fundamental frequency is significantly lower in both whimpering calls and clucking calls, revealing the impact of vocal cord health on frequency (Binazzi *et al*., 2011). Furthermore, differences in peak value and pulse index indicate nonlinear changes in vocal cord vibration, making these features important for distinguishing between normal and abnormal calls (Burkard *et al*., 1984). Finally, the significant increase in spectral line width in whimpering calls further demonstrates the influence of throat mucus on vocal cord vibration characteristics. These changes, analyzed through time-domain and frequency-domain features, effectively differentiate normal calls from abnormal calls, highlighting potential health issues in broiler chickens. As Xu *et al*. (2023) mentioned, animal vocalizations can convey a range of information, and combining acoustic analysis with monitoring chicken calls not only offers significant implications for production performance but also provides new insights into animal welfare. Therefore, focusing on changes in vocalizations during health monitoring can help identify and address health

problems in a timely manner.

Furthermore, our current dataset is sufficient to support the construction of a classification model capable of distinguishing different age-related call types, it must be acknowledged that age-related physiological development significantly affects the acoustic characteristics of the chicken calls. Previous studies have shown that the behavior and call characteristics of chickens change significantly with age and weight increase (Weeks *et al*., 2000), especially in older chickens, where inactivity and lying behaviors increase, which may indirectly affect the production of calls. Additionally, Manteuffel *et al*. (2004) emphasized that calls are closely related to perception and physiological state, meaning that age-related changes in body condition may alter the characteristics of normal calls and distress calls. This is consistent with other research results in the field of bioacoustics and signal processing, where developmental stage and physiological condition have been proven to affect signal characteristics. Our experimental design deliberately integrated voice samples from different age groups (from 2 weeks to 5 weeks) in a balanced distribution. This method ensures that the features learned by the classifier are general across developmental stages, rather than being biased towards specific age-related features. By exposing the model to various acoustic changes related to age, we enhanced its ability to detect pathological patterns (such as patterns caused by laryngeal mucus), which, despite potential age effects, are still distinguishable. We clearly state that a detailed study on the interaction between age and abnormal call patterns is beyond the scope of this research. As Weeks *et al*. (2000) pointed out, studies on the behavior of chickens require strictly controlled conditions and a large sample size to distinguish these complex effects. Therefore, building upon our approach that controls for age variation, we suggest that future research should incorporate longitudinal monitoring and age-stratified analysis to elucidate the developmental trajectory of both normal and abnormal vocalizations across the complete growth cycle. Such investigations would not only refine early warning systems but also provide fundamental insights into the ontogeny of avian vocal patterns in both healthy and pathological states.

Beyond the specific acoustic correlations revealed in this work, our methodological framework demonstrates distinct advantages over prior studies on poultry vocal recognition and health monitoring. Whereas previous research primarily emphasized algorithmic innovation or empirical feature aggregation, our study integrates rigorous statistical feature evaluation with

machine learning classification, yielding both high performance and physiological interpretability. Unlike the chicken voice recognition method by Cheng and Zhong (2015) based on orthogonal matching pursuit (OMP), which reconstructed signals via sparse representation before extracting MFCC, LPC, and PNCC features for SVM classification, our work moves beyond generic signal reconstruction to systematically evaluate which acoustic features are most discriminative. While their sparse representation pipeline improved recognition under noisy environments, feature selection remained empirical, lacking statistical verification of inter-condition differences. In contrast, our independent-samples t-test analysis identified physiologically meaningful features- such as maximum amplitude, effective amplitude, fundamental frequency, and pulse index- that directly reflect pathological changes in broiler vocalizations. This principled approach enabled our SVM classifier to achieve 97.8% accuracy, far exceeding the performance of MFCC-only (77.54%) or LPCC-only (58.89%) models and improving upon the results of Cheng and Zhong's sparse feature method.

Similarly, the improved online multiple kernel classification (OMKC) algorithm developed by Cheng (2014) and colleagues focused on kernel optimization to reduce classification error across generic datasets. Although kernel learning improves algorithmic flexibility, it does not inherently ensure that the selected input features are biologically or statistically meaningful for broiler health assessment. By contrast, our feature evaluation process is hypothesis-driven rather than purely model-driven, ensuring that the selected features are not only computationally efficient but also interpretable in terms of broiler physiology, thereby enhancing both scientific credibility and practical applicability.

The broiler sound recognition framework of Tao *et al*. (2022) and Sun *et al*. (2024) represents another line of research that constructed high-dimensional (60D) multi-domain feature sets and used random forest or kNN to rank feature importance. While these methods achieved competitive classification results (up to 94.16% accuracy in Tao *et al*. and 91.14% in Sun *et al*.), their feature selection relied on post hoc model importance rather than pre-validated statistical differentiation between health conditions. In contrast, our compact t-test-selected feature set avoids overfitting risks inherent to high-dimensional data and simultaneously achieved 98.76% accuracy with random forest, surpassing prior results while reducing computational burden. Moreover, our method directly targets pathological vocal characteristics (such as decreased

amplitude and lowered fundamental frequency caused by respiratory distress) rather than relying on generic acoustic descriptors.

Finally, while Sun *et al*. (2024) advanced broiler health monitoring by integrating sound classification into a cough-rate estimation framework with a visualization platform, their emphasis remained on engineering system deployment rather than on feature interpretability. Our study complements and extends this line of work by providing a statistically grounded, physiologically relevant feature selection strategy that can be seamlessly integrated into such platforms, improving both early disease detection and diagnostic transparency.

In summary, this study presents a transparent and domain-optimized framework for broiler health monitoring by integrating rigorous statistical feature evaluation with advanced machine learning classification. Unlike approaches that rely on empirical feature aggregation, our method employs t-test-based feature selection to ensure that the extracted acoustic parameters are biologically meaningful and closely linked to pathological vocal changes. By focusing on statistically validated, physiologically interpretable features, the proposed system achieves superior recognition accuracy with a compact feature set, effectively avoiding the overfitting issues commonly observed in high-dimensional models. Furthermore, the framework directly targets pathological variations in broiler vocalizations, such as reductions in amplitude and fundamental frequency, rather than relying solely on generic acoustic descriptors, thereby enhancing its diagnostic relevance. Designed for seamless integration with existing signal-processing and monitoring platforms, this approach combines scientific rigor with practical applicability, offering a robust solution for automated broiler health assessment. By moving beyond traditional feature sets, sparse reconstruction techniques, and purely kernel-driven optimization strategies, this research establishes a physiologically grounded, computationally efficient, and high-performance methodology that advances the state of the art in poultry vocal analysis and health monitoring.

Conclusions

This study establishes a rigorous and physiologically interpretable framework for broiler health monitoring through detailed acoustic feature analysis combined with machine learning classification. Unlike prior work that relied on empirical feature aggregation, our approach systematically identifies statistically significant features -including but not limited to maximum

amplitude, effective amplitude, fundamental frequency, and pulse index- that are closely associated with pathological changes in vocalization caused by respiratory disorders. These features reflect how non-healthy physiological states alter vocal cord vibration characteristics, enabling reliable differentiation between normal and abnormal calls.

The framework was validated on vocalization data collected from a batch of 2-week-old white-feathered broilers over a continuous 21-day period, demonstrating high classification accuracy using both support vector machine (97.8%) and random forest (98.76%) models. The use of a statistically validated, multidimensional feature set minimizes overfitting risks while maintaining strong discriminative power across different call types. Importantly, the identified features are physiologically relevant rather than being generic acoustic descriptors, ensuring that the classification results directly reflect health-related vocal changes rather than incidental sound variations.

From an application perspective, this methodology provides a scientifically grounded tool for poultry farms to detect respiratory problems early and objectively, replacing reliance on subjective observation or delayed clinical diagnosis. Since the feature set is compact and computationally tractable, it can be readily integrated into automated monitoring platforms using standard signal processing hardware. By linking acoustic indicators to specific health states, this study enhances both the reliability of early warning systems and the feasibility of large-scale deployment in precision poultry farming.

References

Abdul, Z.K., Al-Talabani, A.K. 2022. Mel frequency cepstral coefficient and its applications: A review. IEEE Access 10:122136-122158.

Akçay, M.B., Oğuz, K. 2020. Speech emotion recognition: Emotional models, databases, features, preprocessing methods, supporting modalities, and classifiers. Speech Commun. 116:56-76.

Balaji, T.K., Annavarapu, C.S.R., Bablani, A. 2021. Machine learning algorithms for social media analysis: A survey. Comput. Sci. Rev. 40:100395.

Balaji, V.R., Maheswaran, S., Babu, M.R., Kowsigan, M., Prabhu, E., Venkatachalam, K. 2020. Combining statistical models using modified spectral subtraction method for embedded system. Microprocess. Microsyst. 73:102957.

Bhardwaj, V., Ben Othman, M.T., Kukreja, V., Belkhier, Y., Bajaj, M., Goud, B.S., Hamam, H. 2022. Automatic speech recognition (ASR) systems for children: A systematic literature review. Appl. Sci. 12:4419.

Biocca, M., Gallo, P., Di Loreto, G., Imperi, G., Pochi, D., Fornaciari, L. 2019. Noise attenuation

provided by hedges. J. Agric. Eng. 50:889.
Binazzi, B., Lanini, B., Romagnoli, I., Garuglieri, S., Stendardi, L., Bianchi, R., Scano, G. 2011. Dyspnea during speech in chronic obstructive pulmonary disease patients: Effects of pulmonary rehabilitation. Respiration 81:379-385.
Bu, Y.F., Qi, X.J., Wen, J.B., Xu, Z.C. 2016. Analysis of acoustic characteristics of 7 forest pest species. J. Nanjing Forest. Univ. Nat. Sci. Ed. 40:179-184.
Burkard, R. 1984. Sound pressure level measurement and spectral analysis of brief acoustic transients. Electroencephalogr. Clin. Neurophysiol. 57:83-91.
Catania, P., Vallone, M. 2013. Noise levels of a track-laying tractor during field operations in the vineyard. J. Agric. Eng. 44:e154.
Cheng, B., Zhong, S. 2014. An improved algorithm of OMKC based on the optimized perceptron with the best kernel. Proc. 2nd Int. Conf. Information Technology and Electronic Commerce, Dalian. pp. 247-251.
Cheng, B., Zhong, S. 2015. A novel chicken voice recognition method using the orthogonal matching pursuit algorithm. Proc. 8th Int. Congr. Image and Signal Processing (CISP), Shenyang. pp. 1266-1271.
Cuan, K., Zhang, T., Huang, J., Fang, C., Guan, Y. 2020. Detection of avian influenza-infected chickens based on a chicken sound convolutional neural network. Comput. Electron. Agr. 178:105688.
Curtin, R.R., Daley, W., Anderson, D.V. 2014. Classifying broiler chicken condition using audio data. Proc. IEEE Global Conf. Signal and Information Processing (GlobalSIP), Atlanta. pp. 1141-1144.
de Carvalho Soster, P., Grzywalski, T., Hou, Y., Thomas, P., Dedeurwaerder, A., De Gussem, M., et al. 2025. Automated detection of broiler vocalizations a machine learning approach for broiler chicken vocalization monitoring. Poultry Sci. 104:104962.
Dong, X., Yin, B., Cong, Y., Du, Z., Huang, X. 2020. Environment sound event classification with a two-stream convolutional neural network. IEEE Access 8:125714-125721.
Du, X., Teng, G., Wang, C., Carpentier, L., Norton, T. 2021. A tristimulus-formant model for automatic recognition of call types of laying hens. Comput. Electron. Agr. 187:106221.
Hegde, J., Shenoy, M. 2024. Analysis of frequencies and pitches for vocal cord paralysis classification through transfer learning. Proc. IEEE Int. Conf. Electronics, Computing and Communication Technologies (CONECCT), Bangalore. pp. 1-6.
Huang, J., Zhang, T., Cuan, K., Fang, C. 2021. An intelligent method for detecting poultry eating behaviour based on vocalization signals. Comput. Electron. Agr. 180:105884.
Li, G., Hou, Q., Zhang, C., Jiang, Z., Gong, S. 2021. Acoustic parameters for the evaluation of voice quality in patients with voice disorders. Ann. Palliat. Med. 10:13036-13136.
Li, Z., Zhang, T., Cuan, K., Fang, C., Zhao, H., Guan, C., Qu, H. 2022. Sex detection of chicks based on audio technology and deep learning methods. Animals 12:3106.
Lin, J. 2001. Feature extraction of machine sound using wavelet and its application in fault diagnosis. NDT & E Int. 34:25-30.
Mahdavian, A., Minaei, S., Marchetto, P.M., Almasganj, F., Rahimi, S., Yang, C. 2021. Acoustic features of vocalization signal in poultry health monitoring. Appl. Acoust. 175:107756.
Manteuffel, G., Puppe, B., Schön, P.C., 2004. Vocalization of farm animals as a measure of welfare. Appl. Anim. Behav. Sci. 88:163-182.

Moura, D.J.D., Nääs, I.D.A., Alves, E.C.D.S., Carvalho, T.M.R.D., Vale, M.M.D., Lima, K.A.O.D. 2008. [Noise analysis to evaluate chick thermal comfort].[Article in Portuguese with English abstract]. Sci. Agric. 65:438-443.

Peters, G., Wendler, O., Böhringer, D., Gostian, A.O., Müller, S.K., Canziani, H., Hesse, N. et al. 2021. Viscoelasticity of human laryngeal mucus from the vocal folds. Appl. Sci. (Basel) 11:3011.

Sun, Y.M., Liu, F., Chen, L.S. 2019. A comparative study on the song structure of three grasshopper species. J. Shihezi Forest. Univ. Nat. Sci. Ed. 37:565-569.

Sun, Z., Tao, W., Gao, M., Zhang, M., Song, S., Wang, G., 2024. Broiler health monitoring technology based on sound features and random forest. Eng. Appl. Artif. Intell. 135:108849.

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

Weeks, C.A., Danbury, T.D., Davies, H.C., Hunt, P., Kestin, S.C., 2000. The behaviour of broiler chickens and its modification by lameness. Appl. Anim. Behav. Sci. 67:111-125.

Wei, J.R., Zhu, L.X., Zhang, S.H., Chen, P.L. 2023. Research progress of livestock sound monitoring technology. Mod. Agric. Equip.44:8-15.

Yadav, S., Keerthana, M., Gope, D., Ghosh, P.K. 2020. Analysis of acoustic features for speech sound based classification of asthmatic and healthy subjects. Proc. IEEE Int. Conf. Acoustics, Speech and Signal Processing (ICASSP), Barcelona. pp. 6789-6793.

Yao, W., Deng, W., Xu, Y., Sun, W., Liu, L., Shen, M. 2023. Poultry sub-health status monitoring and health warning prospect. J. Nanjing Agric. Univ. 46:635-644.

Yin, J.J., Li, W.G., Liu, Y.F., Xiao, D.Q. 2024. Sex identification of ducklings based on acoustic signals. Poultry Sci. 103:103711.

Yu, C., Kang, M., Chen, Y., Wu, J., Zhao, X. 2020. Acoustic modeling based on deep learning for low-resource speech recognition: An overview. IEEE Access 8:163829-163843.

Yu, Z., Guindani, M., Grieco, S.F., Chen, L., Holmes, T.C., Xu, X. 2022. Beyond t test and ANOVA: Applications of mixed-effects models for more rigorous statistical analysis in neuroscience research. Neuron 110:21-35.

Xu, C., Rao, W., Chng, E.S., Li, H. 2020. SPEX: Multi-scale time domain speaker extraction network. IEEE/ACM Trans. Audio Speech Lang. Process. 28:1370-1384.

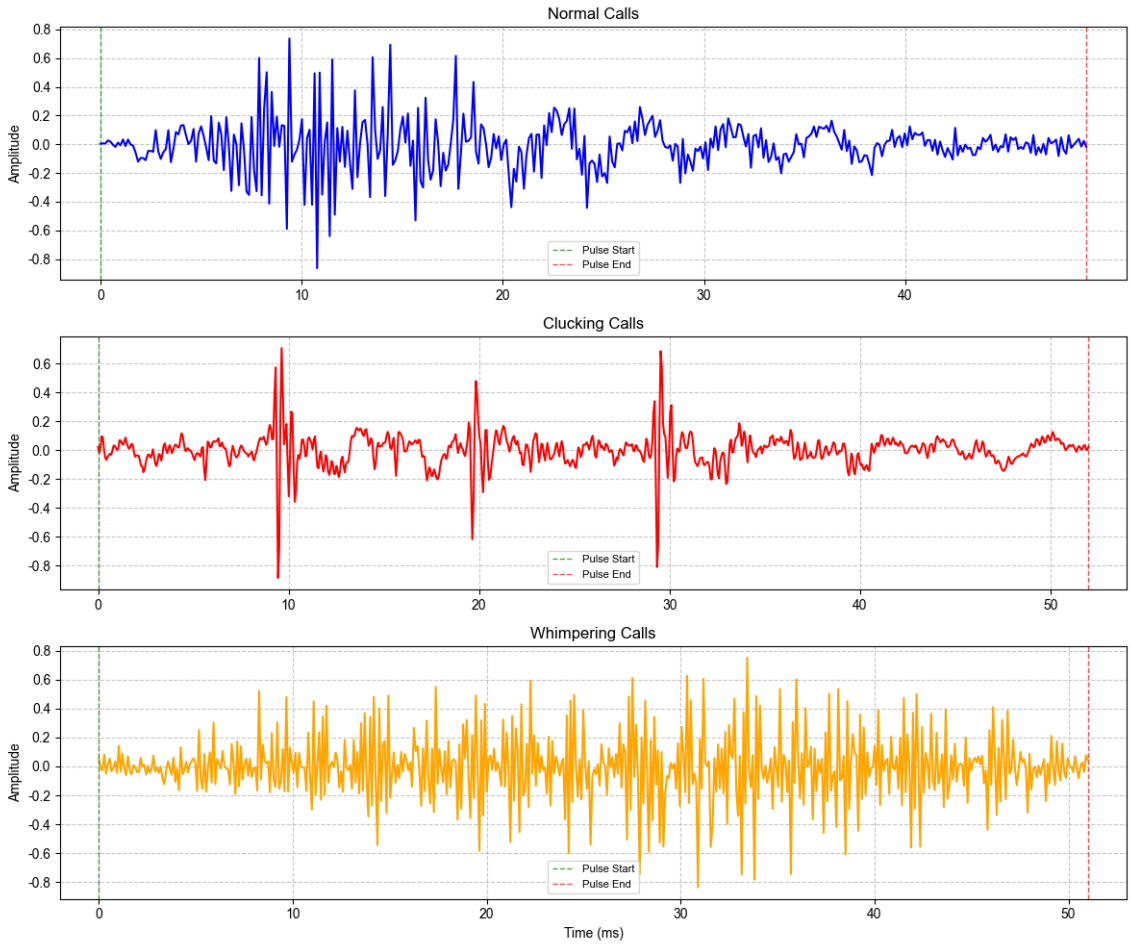

Figure 1. Frame segmentation of vocal signals.

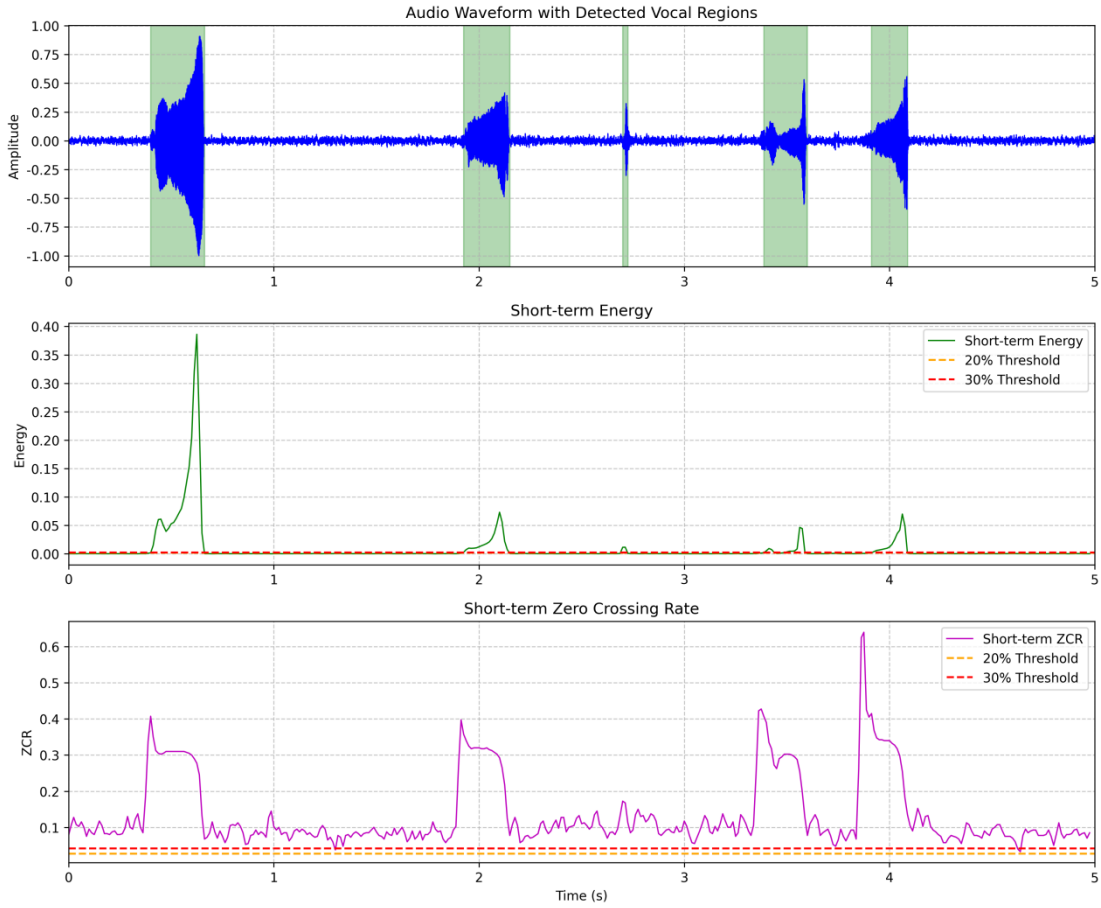

Figure 2. Effect diagram of the endpoint detection process for chicken crowing sounds.

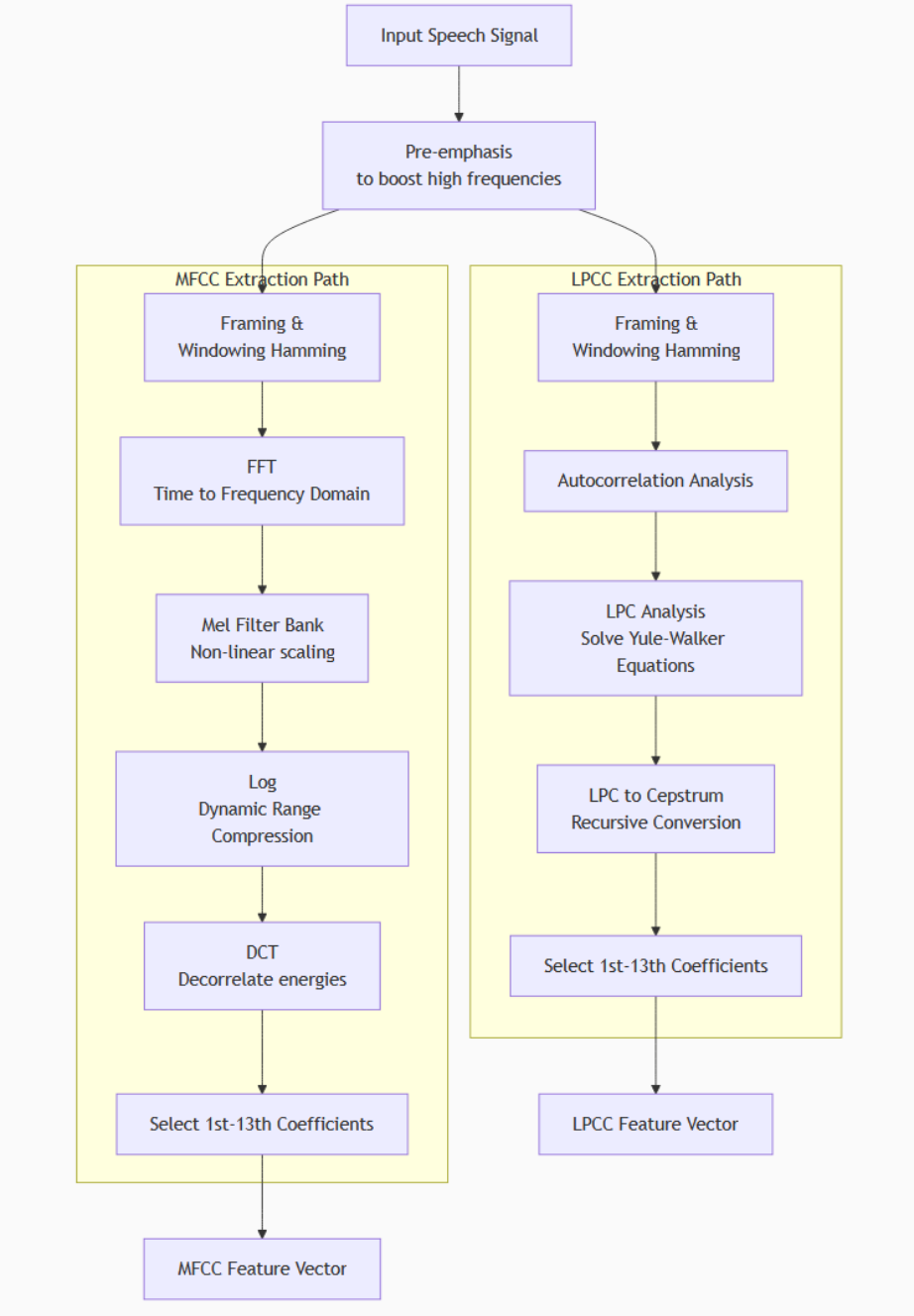

Figure 3. Extraction process of mel-frequency cepstral coefficients (MFCC) and linear prediction cepstral coefficients.

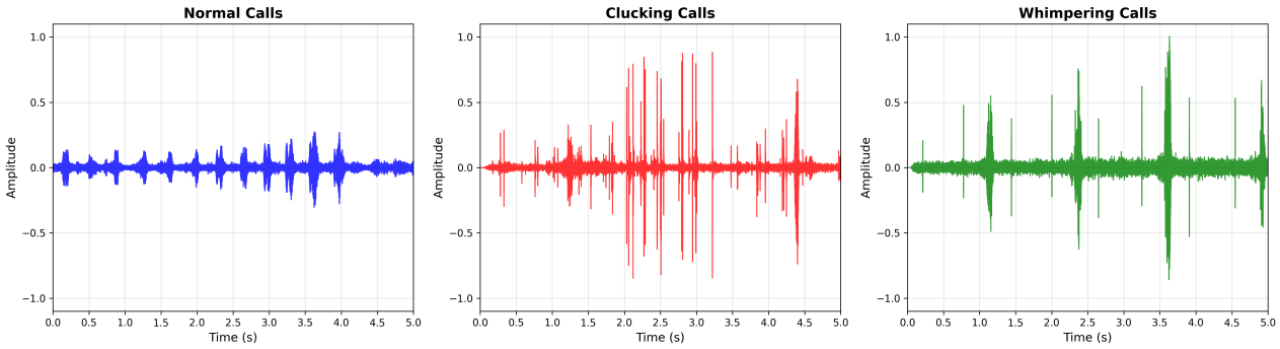

Figure 4. Waveform diagram of "normal calls", "clucking calls" and "whimpering calls".

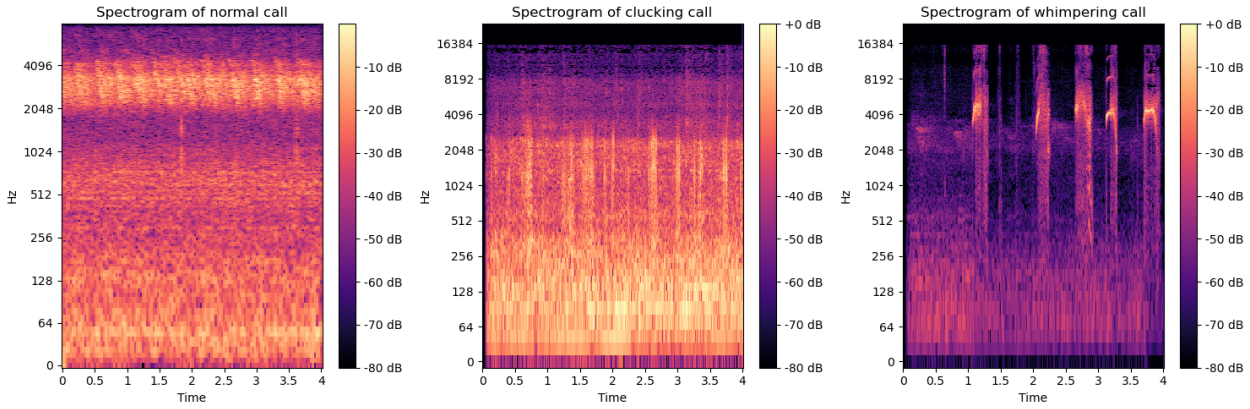

Figure 5. Spectrograms of the "normal calls", "clucking calls" and "whimpering calls".

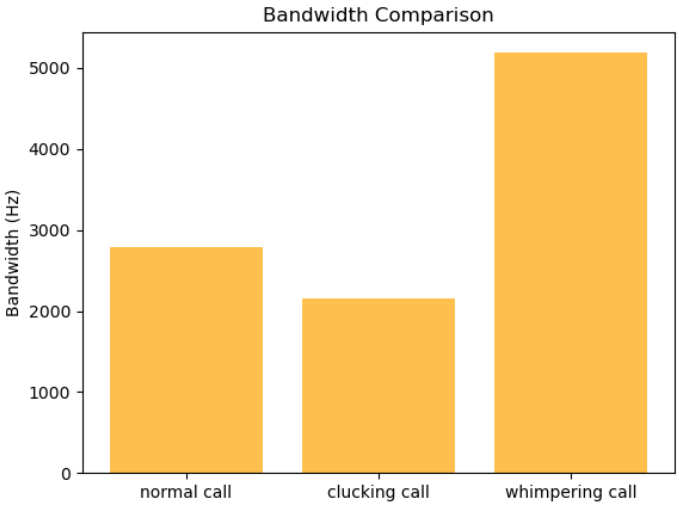

Figure 6. The bandwidth of the "normal calls", "clucking calls" and "whimpering calls".

Table 1. Time-domain and frequency-domain feature formulations.

Feature category	Feature name	Mathematical formulation	Description
Time-domain	Maximum amplitude	$A_{max} = \max(\lvert x[n] \rvert)\ , n = 0,1,\ldots,N-1$ $x[n]$: The raw amplitude value of the n-th sample in the discrete signal frame. This value can be positive or negative.	The maximum absolute amplitude value within a signal frame, reflecting the instantaneous peak intensity.
	Peak value	$A\_p = \max(x[n]), n = 0,1,\ldots,N-1$ $x[n]$: amplitude at sample n.	The maximum positive amplitude, indicating the instantaneous maximum output.
	Effective amplitude	$A_{RMS} = \mathrm{sqrt}\left(\frac{1}{N} * \mathrm{sum}(x[n]^2)\right)\ , n = 0,1,\ldots,N-1$ $\frac{1}{N} * \mathrm{sum}(x[n]^2)$: The mean (average) of the squared values. This is the mean square value, representing the average power of the signal. $\mathrm{sqrt}\left(\frac{1}{N} * \mathrm{sum}(x[n]^2)\right)$: The square root of the mean square value. This returns the value to the original unit of amplitude, giving the RMS amplitude.	A_{RMS} represents the average power or overall energy content of the signal frame and serves as a robust measure of loudness.
	Variance	$\sigma^2 = \frac{1}{N}\sum_{n=0}^{N-1}(x[n]-\mu)^2$ $\mu = \frac{1}{N}\sum_{n=0}^{N-1}x[n]$ $n = 0,1,\ldots,N-1$ $\sum_{n=0}^{N-1}(x[n]-\mu)^2$: The sum of squared deviations. This adds up all the individual squared differences. $\frac{1}{N}\sum_{n=0}^{N-1}(x[n]-\mu)^2$: The mean of the squared deviations. μ:The mean amplitude of the signal frame, indicating the variability of the amplitude distribution.	It measures the deviation of the signal from its mean value, indicating the variability of the amplitude distribution.
	Skewness	$\gamma = ((1/N)\,\Sigma\,(x[n]-\mu)^3) / \sigma^3$ $n = 0,1,\ldots,N-1$ $x[n]$: The amplitude of the n-th sample in the	It measures the asymmetry of the amplitude

		discrete signal frame (can be positive or negative). N: The total number of samples in the frame. μ: The mean amplitude of the signal frame, representing the DC offset or central value of the signal.	distribution; $\gamma > 0$ indicates a right-skewed distribution, while $\gamma < 0$ indicates a left-skewed distribution.
	Peak-to-Peak Value	$$V_{pp} = \max(\lvert x[n]\rvert) - \min(\lvert x[n]\rvert)$$ $$n = 0,1, \dots, N-1$$ $\max(\lvert x[n]\rvert)$: The maximum function applied to the raw signal values. It returns the highest (most positive) amplitude value in the frame. $\min(\lvert x[n]\rvert)$: The minimum function applied to the raw signal values. It returns the lowest (most negative) amplitude value in the frame.	The difference between the maximum and minimum amplitude values in the frame, representing the total dynamic range.
	Kurtosis	$$\beta = \frac{\frac{1}{N}\sum_{n=0}^{N-1}(x[n]-\mu)^4}{\left(\frac{1}{N}\sum_{n=0}^{N-1}(x[n]-\mu)^2\right)^2}$$ $$n = 0,1,\dots,N-1$$ $\frac{1}{N}\sum_{n=0}^{N-1}(x[n]-\mu)^4$: The mean of the fourth-power deviations. This is the numerator. $\left(\frac{1}{N}\sum_{n=0}^{N-1}(x[n]-\mu)^2\right)^2$:The square of the variance (σ^2). This is the denominator.	Measures the "tailedness" of the amplitude distribution relative to a normal distribution. Higher values indicate heavier tails and a sharper peak.
	Crest factor	$$C = \frac{\max(\lvert x[n]\rvert)}{A_{RMS}}$$ $$n = 0,1,\dots,N-1$$ $\max(\lvert x[n]\rvert)$: The maximum absolute amplitude of the signal frame. A_{RMS}: Root Mean Square (RMS) amplitude of the frame, representing the average signal magnitude considering both positive and negative values.	The ratio of the signal's peak amplitude to its RMS amplitude, indicating the presence of extreme peaks relative to the average signal level.
	Margin index	$$L = \frac{A\_p}{\left(\frac{1}{N}\sum_{n=0}^{N-1}\sqrt{\lvert x[n]\rvert}\right)^2}$$ $$n = 0,1,\dots,N-1$$ A_p : The peak amplitude. $\left(\frac{1}{N}\sum_{n=0}^{N-1}\sqrt{\lvert x[n]\rvert}\right)^2$: The square of that average	A measure similar to the crest factor but more sensitive to impulse signals, calculated as the ratio of peak amplitude to the

		value. This is the denominator.	square of the averaged square root of absolute amplitude.
	Pulse index	$$\mathrm{PI} = \frac{\max(\lvert x[n] \rvert)}{(\frac{1}{N} \Sigma \lvert x[n] \rvert)}$$ $$n = 0,1,\ldots,N-1$$ $\max(\lvert x[n] \rvert)$: The maximum absolute amplitude in the signal frame (peak value). $\frac{1}{N} \Sigma \lvert x[n] \rvert$: The mean absolute amplitude, representing the overall average signal intensity.	It represents the strength of impulsive components in the signal.
	Waveform index	$$\mathrm{WI} = \frac{\mathrm{sqrt}(\frac{1}{N} \Sigma x[n]^2)}{(\frac{1}{N} \Sigma \lvert x[n] \rvert)}$$ $$n = 0,1,\ldots,N-1$$ $\mathrm{sqrt}(\frac{1}{N} \Sigma x[n]^2)$:The root mean square (RMS) amplitude of the signal, representing its average power. $\frac{1}{N} \Sigma \lvert x[n] \rvert$: The mean absolute amplitude of the signal.	It measures the waveform's similarity to a sinusoid (approximately 1.11 for a sine wave and 1.0 for a square wave).
Frequency-domain	Fundamental Frequency (F0)	$$F0 = f_k$$ where k is the index of the maximum magnitude in the spectrum within a plausible frequency range.	$F0$ is the lowest frequency of vibration of the vocal cords, perceived as the pitch of the call. It is found by identifying the strongest harmonic in the spectrum.
	Spectral centroid	$$C = \frac{\mathrm{sum}(f_k * \lvert x[k] \rvert)}{\mathrm{sum}(\lvert x[k] \rvert)}$$ $$k = 0,1,\ldots,K-1$$ $\mathrm{sum}(f_k * \lvert x[k] \rvert)$: The sum of each frequency multiplied by its energy. This is the numerator. $\mathrm{sum}(\lvert x[k] \rvert)$: The total energy (sum of all magnitudes) in the spectrum. This is the denominator. $x[k]$: $x[k]$ is its discrete Fourier transform	The spectral centroid is the "center of mass" or weighted average of the frequencies present in the spectrum. A higher centroid indicates a brighter sound

		(DFT), defined over frequency bins k=0,1,…,K-1	with more high-frequency energy.
	Frequency RMS	$$f\_RMS = \frac{\text{sqrt}(\ (\Sigma\ f\_k^2\ \|X[k]\|)}{(\Sigma\ \|X[k]\|)\ )}$$ $$k = 0,1,\dots,K-1$$ f_k: The frequency value corresponding to the KTH spectral line (unit: Hz). X[k] : The amplitude of the KTH spectral line after the discrete Fourier transform (DFT) (usually a linear value, not dB).	It represents the RMS frequency distribution, describing the spread of energy across frequencies.
	Spectral bandwidth	$$B = \text{sqrt}\left(\frac{\text{sum}((f_k - c)^2 * \|x[k]\|)}{\text{sum}(\|x[k]\|)}\right)$$ $$k = 0,1,\dots,K-1$$ $\text{sum}((f_k - c)^2 * \|x[k]\|)$: The energy-weighted sum of the squared deviations. $\text{sum}(\|x[k]\|)$: The total energy in the spectrum.	Spectral bandwidth measures the spread of the spectrum around the centroid. A high bandwidth means the spectrum is wide and contains a broad range of frequencies. A low bandwidth means the spectral energy is concentrated tightly around the centroid.
	Spectral line width	$$\Delta f = f\_high - f\_low\ (half - power\ bandwidth)$$ f_high and f_low: upper and lower cut-off frequencies. These two frequency points are defined on the power spectral density (PSD) curve of the spectrum.	The spectral line width indicates the degree of spectral concentration.
	Spectral rolloff	Spectral rolloff is the frequency R below which a certain percentage γ of the total spectral energy is contained.	Spectral rolloff is a measure of the skewness of the spectral shape.

Table2. Comparison of MFCC and LPCC features.

Feature	Primary domain	Underlying principle	Key applications
MFCC	Cepstrum (perceptual)	Mimics the non-linear frequency resolution of the human ear (Mel scale).	Speech recognition, sound classification, music information retrieval.
LPCC	Cepstrum (physical)	Models the acoustic theory of speech production (vocal tract source-filter).	Speaker recognition, speech synthesis, voice pathology detection.

Table 3. Independent samples *t*-test of acoustic features between normal calls and other call types.

Acoustic feature	Normal calls	Whimpering calls	t(Whimper)	p(Whimper)	Clucking calls	*t*-value	p-value
Maximum amplitude	0.685±0.113	0.503±0.094	-28.03	<0.001	0.802±0.138	14.813127	<0.001
Effective amplitude	0.108±0.011	0.030±0.002	-154.55	0.000	0.070±0.012	-52.427058	<0.001
Variance	0.012±0.002	0.001±0.000	-101.50	0.000	0.005±0.002	-51.116122	<0.001
Peak-to-peak value	1.329±0.211	0.887±0.192	-35.00	<0.001	1.400±0.234	5.137306	<0.001
Kurtosis	0.001±0.000	0.000±0.000	-46.95	<0.001	0.001±0.000	-18.827250	<0.001
Peak value	6.356±0.827	16.701±2.659	84.06	0.000	11.659±1.911	57.616207	<0.001
Margin index	6.356±0.827	16.701±2.659	84.06	0.000	11.659±1.911	57.616207	<0.001
Skewness	-5.741±42.015	-9223.372±4779.168	-122.55	0.000	947.884±5043.693	-4.226571	<0.001
Pulse index	9.707±1.545	26.579±4.686	77.38	0.000	20.893±3.744	62.486304	<0.001
Waveform index	1.522±0.064	1.588±0.042	19.22	<0.001	1.789±0.075	60.905236	0.000
Fundamental frequency (Hz)	501.395±319.346	145.674±0.280	-25.20	<0.001	146.844±1.398	-25.121653	<0.001
First resonance peak (Hz)	0.003±0.011	0.001±0.006	-3.58	<0.001	0.001±0.010	-6.591302	<0.001
Second resonance peak (Hz)	0.003±0.010	0.001±0.007	-3.31	<0.001	0.001±0.011	-6.793243	<0.001
Third resonance peak (Hz)	0.003±0.009	0.001±0.007	-2.53	0.011	0.001±0.011	-6.717301	<0.001
Spectral line width (Hz)	0.026±0.004	0.070±0.001	227.65	0.000	0.074±0.003	221.910527	0.000
Frequency root mean square (Hz)	20.080±2.728	8.748±0.559	-92.09	0.000	12.663±2.899	-42.155079	<0.001

Table 4. Machine learning model performance evaluation with different feature sets.

Model	Feature set	Accuracy	Precision	Recall	F1 score
SVM	*t*-test selected	0.978	0.978	0.979	0.978
	MFCC	0.775	0.777	0.776	0.775
	LPCC	0.589	0.583	0.592	0.579
	MFCC+LPCC	0.783	0.784	0.783	0.782
Naive bayes	*t*-test selected	0.878	0.884	0.876	0.874
	MFCC	0.548	0.723	0.556	0.528
	LPCC	0.616	0.609	0.611	0.609
	MFCC+LPCC	0.548	0.723	0.556	0.528
KNN	*t*-test selected	0.957	0.958	0.956	0.956
	MFCC	0.537	0.544	0.532	0.520
	LPCC	0.378	0.400	0.388	0.318
	MFCC+LPCC	0.534	0.539	0.530	0.525
Random forest	*t*-test selected	0.988	0.988	0.988	0.988
	MFCC	0.778	0.780	0.778	0.777
	LPCC	0.782	0.789	0.779	0.780
	MFCC+LPCC	0.813	0.814	0.811	0.810