Research Article

Early Distress Detection in Etawah Crossbreed Goats: A Pre-Processing Pipeline

Zia ul Rahman Fithron[1], Kuswati[1], Barlian Henryranu Prasetio[2], Muhammad Halim Natsir[1], Lilik Eka Radiati[1], Marjuki[1], Ahmad Khoirul Umam[1], Muhammad Helmi[1], Putu Wahyu Kusuma Wardhana[1], Tri Eko Susilorini[1]*

[1]Faculty of Animal Science and Technology, Universitas Brawijaya, Malang, Indonesia; [2]Faculty of Computer Science, Universitas Brawijaya, Malang, Indonesia.

Abstract | Vocalization analysis offers a non-invasive approach to livestock welfare monitoring, yet automated detection of goat distress calls remains challenging due to noise interference in outdoor farm environments. This study presents and evaluates a multi-stage signal pre-processing pipeline applied to a field-recorded distress sequence from an adult female Etawah Crossbreed goat (duration: 104.70 s; 16,000 Hz; noise floor: −18.1 dB RMS) captured during a naturally occurring acute-pain event. The pipeline integrates noise reduction (Wiener Filtering, and MMSE-STSA), energy-based Voice Activity Detection (VAD) with a −30 dB threshold and 150-ms merge gap, adaptive Hamming-windowed segmentation (10-40 ms frames) for dynamic signal pre-conditioning, and peak normalization, followed by fixed-window (32 ms) extraction of Mel-Frequency Cepstral Coefficients (MFCCs), fundamental frequency (F0), spectral centroid, spectral bandwidth, and zero-crossing rate (ZCR). The Wiener filter achieved the best performance (SNR improvement +10.3 dB; Log-Spectral Distance 1.12 dB) while preserving biological harmonic structures. The VAD isolated 29 discrete vocalization events (signal occupancy 34.6%; recall 93.3%, precision 96.6%, F1 94.9%), with call durations of 0.22–2.29 s (1.25 ± 0.70 s) classified into short, medium, and sustained calls. Feature analysis yielded a mean F0 of 314.1 ± 160.3 Hz, spectral centroid of 1,760.8 ± 510.8 Hz, and ZCR of 0.123 ± 0.043, with the spectral centroid proving highly discriminative across call classes. The pipeline effectively prepares raw livestock recordings for downstream machine learning, supporting automated early distress detection in precision livestock farming.

Keywords | Early distress detection, Etawah crossbreed vocalization, MFCC, MMSE-STSA, Voice activity detection, Wiener filter

Received | June 03, 2026; **Accepted** | August 01, 2026; **Published** | August 29, 2026
***Correspondence** | Tri Eko Susilorini, Faculty of Animal Science and Technology, Universitas Brawijaya, Malang, Indonesia; **Email:** triekos@ub.ac.id
Citation | Fithron ZR, Kuswati, Prasetio BH, Natsir MH, Radiati LE, Marjuki, Umam AK, Helmi M, Wardhana PWK, Susilorini TE (2026). Early distress detection in etawah crossbreed goats: A pre-processing pipeline. Adv. Anim. Vet. Sci., 14(9):2108-2121.
DOI | https://dx.doi.org/10.17582/journal.aavs/2026/14.9.2108.2121
ISSN (Online) | 2307-8316

INTRODUCTION

The Etawah Crossbreed goat (Locally Peranakan Etawah, PE) is among the most widely kept and economically important goat types in Indonesia. It originated from crosses between Jamnapari goats, introduced from the Etawah region of Uttar Pradesh, India, during the colonial era, and indigenous Kacang goats, producing a dual-purpose animal valued for both milk and meat that is better adapted to local tropical conditions than its Jamnapari ancestor from its breeding center in Central Java, the PE goat has dispersed across Java and beyond and is now reared throughout the country, predominantly by smallholders keeping only a few head each, for whom it represents an accessible source of household income and high-value goat milk (Anggraeni

et al., 2020). This combination of national prevalence, smallholder dependence, and dual-purpose value makes the welfare and health of PE goats a matter of direct economic and food-security relevance, and motivates the present focus on this breed (Susilorini *et al*., 2014).

Early detection of physiological distress remains a critical unresolved problem in precision livestock farming (PLF): In goats, prolonged undetected distress drives immunosuppression, reduced intake, and impaired growth, and subclinical conditions that surface as behavioral and vocal change before overt clinical signs cause disproportionate losses (Manteuffel *et al*., 2004). Conventional visual monitoring is labor-intensive, subjective, and unsuited to farms where one operator manages hundreds of animals and early distress goes unseen at night motivating automated, continuous alternatives (Manikandan and Neethirajan, 2025).

Among distress indicators, vocalization is uniquely non-invasive and remotely detectable, requiring no contact sensors (Manteuffel *et al*., 2004; Briefer, 2012). Distress and pain of goat calls show elevated fundamental frequency (*F0*), higher call rate, and altered harmonic structure relative to neutral states, demonstrated in goats themselves rather than inferred from other species (Briefer and McElligott, 2012). That such vocal change can anticipate clinical signs to established mainly in other livestock; the often-cited hours-to-days lead time derives from cattle and the equivalent lead time has not been quantified in goats (Gavojdian et al., 2024). Therefore treat early warning as a working hypothesis grounded in the documented arousal acoustic coupling in caprines rather than as a goat-specific fact, and this study does not test it: it addresses the prerequisite signal-processing problem that must be solved first (Manteuffel *et al*., 2004).

The goat-specific acoustic evidence base, however, was built almost entirely under controlled conditions: The reference characterizations of Briefer and McElligott (2012) and Briefer (2012) recorded calls at close range against quiet backgrounds at high signal-to-noise ratios, where F0, harmonic structure, and spectral parameters are measured directly. A working-farm recording differs from these references in measurement rather than biology, the vocal source is unchanged, but what reaches a single uncalibrated microphone is not. Wideband farm noise overlaps spectrally with the call and biases precisely the features those studies report: spectral centroid and bandwidth are pulled upward by broadband energy, ZCR is inflated by high-frequency interference, and pitch trackers readily lock onto low-frequency structural artifacts as spurious F0. Laboratory acoustic values therefore cannot be transferred to farm audio without first restoring signal quality, which is why a dedicated pre-processing stage, rather than direct feature extraction, is the prerequisite for any comparison with the controlled-condition literature. The absence of a standardized, validated pre-processing pipeline for goat vocalizations in farm conditions thus represents a critical methodological gap.

Mel-Frequency Cepstral Coefficients (MFCCs), the established representation for affective and species-level information in animal calls (Davis and Mermelstein, 1980) lose discriminative value sharply when the input is noisy, imprecisely segmented, or poorly normalized pre-processing quality thus sets a ceiling on every downstream feature.

A deployable pipeline must integrate four sequential functions: noise reduction, voice activity detection (VAD), adaptive segmentation, and feature extraction. Because no public corpus of farm-recorded, expert-annotated goat distress calls exists, an end-to-end comparison against a prior goat pipeline is not yet possible; we therefore benchmark each stage against its own established reference rather than leave it unvalidated. The three noise-reduction methods are compared head-to-head (spectral subtraction, Wiener filtering, and the MMSE-STSA estimator; Ephraim and Malah, 1984) on quantitative criteria (SNR gain, log-spectral distance), and VAD output is evaluated against a human-expert manual annotation of the full recording, the de facto reference standard for vocalization detection. To our knowledge no integrated, quantitatively evaluated pre-processing pipeline has been reported for goat distress vocalizations under realistic farm conditions; existing work addresses single stages, uses laboratory conditions, or targets other species. We accordingly design, implement, and evaluate this pipeline on a field-recorded goat distress sequence, aiming to establish a reproducible, deployable framework that yields high-quality features from farm audio as the foundation for subsequent early distress detection and welfare monitoring.

MATERIALS AND METHODS

Study design and sample collection

This study utilized an exploratory, proof-of-concept methodological design to validate a bioacoustic pre-processing pipeline under authentic, noisy field conditions. Because the objective was to evaluate the technical robustness of the signal-enhancement framework rather than biological generalizability, the pipeline was tested using a single, high-density continuous field recording. To comply with strict international animal welfare guidelines, vocalizations were captured opportunistically from a naturally occurring incidental event in January 2026 on a smallholder farm in Trenggalek, Indonesia, rather than being experimentally induced.

An acute, painful distress event was opportunistically recorded from an adult female Etawah Crossbreed goat whose hoof became accidentally trapped in wooden slat flooring. Due to ethical constraints regarding induced animal suffering, this high-pain state could not be replicated. Recording was strictly observational and non-invasive: no procedure or restraint was applied to the animal for research purposes, and as soon as the entrapment was identified the hoof was released and the goat was promptly examined and given appropriate husbandry and veterinary care, recovering without lasting harm. The recording was made on privately owned livestock with the prior informed consent of the farm owner. A total of 29 vocalization events were captured across a 104.70 s recording, which featured a high farm background noise floor (-18.1 dB RMS). Acoustic data were acquired at a sampling rate of 16,000 Hz using a portable digital recorder equipped with a built-in omnidirectional microphone (TNW A37, TNW Tech, China) without a windscreen. The signal was converted using FFmpeg (v6.0) into uncompressed 16-bit linear PCM WAV format. The extracted features function strictly as intra-sample descriptive indicators for this specific sequence and do not imply population-level parameters.

Signal pre-processing pipeline

All signal processing was implemented in Python 3.10 using NumPy (v1.24), SciPy (v1.10), and librosa (v0.10) (McFee et al., 2015). The converted waveform was processed through four sequential operations: noise reduction, energy-based voice activity detection, adaptive Hamming-windowed segmentation, and peak amplitude normalization, after which acoustic features were extracted from each isolated vocalization event. Noise reduction was applied with a 25-ms analysis frame and 10-ms hop and a noise estimate derived from the initial 150 ms of the recording; voice activity detection used a −30 dB energy threshold with a minimum call duration of 50 ms, a 150-ms merge gap, and 80-ms symmetric padding; segmentation used adaptive frame sizes of 10–40 ms with 50% overlap and overlap-add reconstruction; and amplitude normalization scaled each segment to a peak of 0.90. An overview of the complete experimental design is presented in Figure 1, and each step is described in detail below.

Noise reduction (Wiener filtering)

Three established noise reduction methods were implemented and compared. All methods shared a common short-time Fourier transform (STFT) framework with a 25-ms Hamming-windowed analysis frame and 10-ms hop interval. A noise power spectral density (PSD) estimate was derived from the initial 150 ms of the recording, a segment verified to contain only background noise prior to any vocalization onset. The frequency-domain Wiener gain is computed as $H(k) = S_x(k) / [S_x(k) + S_n(k)]$, where $S_x(k)$ denotes the estimated signal PSD and $S_n(k)$ the noise PSD at frequency bin k. The gain approaches unity in high-SNR bins and approaches zero in noise-dominated bins, providing frequency-selective suppression without phase distortion (Loizou, 2007).

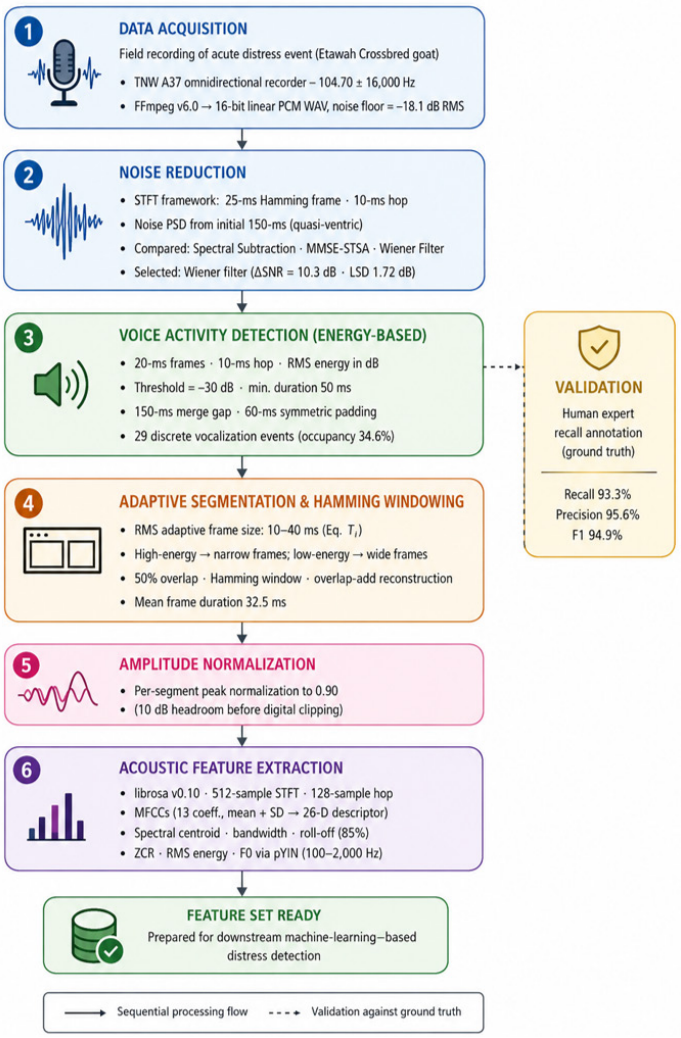

Figure 1: Overview of the multi-stage signal pre-processing pipeline for goat distress vocalizations. Raw field audio (104.70 s, 16,000 Hz) passes sequentially through (1) data acquisition and PCM conversion, (2) noise reduction in which three methods are compared head-to-head and the Wiener filter is selected (ΔSNR +10.3 dB; LSD 1.12 dB), (3) energy-based voice activity detection (−30 dB threshold, 150-ms merge gap, 80-ms padding) isolating 29 discrete vocalization events, (4) RMS-adaptive Hamming-windowed segmentation (10–40 ms frames, 50% overlap, overlap-add reconstruction), (5)per-segment peak amplitude normalization to 0.90, and (6) acoustic feature extraction (MFCCs, spectral centroid, bandwidth, roll-off, ZCR, RMS energy, and pYIN F0). The voice activity detection stage is validated against a human-expert manual annotation serving as the ground-truth reference (recall 93.3%, precision 96.6%, F1 94.9%). Solid arrows denote the sequential processing flow; the dashed connector denotes validation against the ground truth.

The noise power spectral density (PSD) profile was estimated from the initial 150 ms window of the continuous field recording. To ensure this baseline segment contained purely stationary environmental noise and was completely free of low-amplitude call onsets or pre-vocal breath structures, a dual-verification protocol was conducted. Visual inspection of the high-resolution STFT spectrogram confirmed a total absence of harmonic trajectories, formant structures, or pitch tracking below 4,000 Hz. Second, an auditory playback audit verified that the window contained only wideband farm background ambient noise, thereby establishing a clean, non-contaminated noise baseline for the subsequent Wiener filtering stage.

Spectral subtraction

The power spectrum of the estimated noise is subtracted from that of the noisy signal, with a spectral floor parameter $\beta = 0.005$ to prevent negative power values ("musical noise") and an over-subtraction factor $\alpha = 1.5$. The magnitude spectrum of the enhanced frame is reconstructed using the original phase (Boll, 1979).

MMSE-STSA

The Ephraim–Malah estimator minimizes the mean-square error of the short-time spectral amplitude. A priori SNR is tracked using the decision-directed approach with a smoothing factor $\alpha = 0.98$ (Ephraim and Malah, 1984; Loizou, 2007). This formulation produces spectrally smooth estimates with fewer musical noise artifacts than simple subtraction methods.

Voice activity detection

An energy-based VAD was applied to the noise-reduced signal. The signal was divided into 20-ms frames with 10-ms hop. The root-mean-square (RMS) energy of each frame was computed and converted to decibels relative to a full-scale reference. Frames were classified as "active" (vocalization present) if their RMS exceeded a threshold of –30 dB, selected empirically to balance sensitivity with false-positive rate in this outdoor recording environment a threshold 10 dB higher than is commonly employed for human speech VAD (Sohn *et al*., 1999). Active segments separated by gaps of less than 150 ms were merged to prevent fragmentation of continuous calls. A temporal padding of 80 ms was applied symmetrically around each detected segment to capture call onset and offset transients. The minimum accepted call duration was 50 ms. To evaluate the classification and segmentation accuracy of the energy-based VAD algorithm, a retrospective human-in-the-loop validation was implemented to establish a definitive ground truth database. A human bioacoustic expert visually inspected auditorily audited the entire 104.70 s raw audio recording. True vocalizations were differentiated from non-voiced transient biological sounds (e.g., coughs or sneezes) and mechanical farm interferences based on the presence of clear, periodic harmonic trajectories and elevated fundamental frequency structures characteristic of acute pain-induced distress. Aand also to confirm the VAD algorithm, a retrospective verification was conducted against a human-expert manual segmentation baseline

Adaptive segmentation and hamming windowing

Each isolated vocalization event was further divided into overlapping analysis frames using an adaptive frame-size strategy. The local RMS energy of each frame was used to select frame duration: high-energy frames corresponding to consonantal or transient events were analyzed following formula (Mauch and Dixon, 2014):

$$T\_f = T\_max - [(E\_frame - E\_min) / (E\_max - E\_min)] * (T\_max - T\_min)$$

Where; T_min is set to 10 ms and T_max is set to 40 ms. This adaptive boundary ensures that high-energy, rapid call onsets are assigned narrower frames (approaching 10 ms) to maximize temporal resolution, whereas stable, low-energy harmonic regions are assigned wider frames (approaching 40 ms) to enhance spectral resolution prior to Mel-Frequency Cepstral Coefficients (MFCC) extraction. A 50% overlap between consecutive frames was maintained throughout. Each frame was multiplied by a Hamming window $w(n) = 0.54 - 0.46 \cdot \cos(2\pi n/(N-1))$ to reduce spectral leakage (Harris, 1978). Signal reconstruction employed the overlap-add (OLA) method with window normalization to generate a dynamically smoothed waveform prior to final feature extraction (Oppenheim and Schafer, 2009).

Amplitude normalization

Peak normalization was applied to the reconstructed signal to standardize amplitude levels for downstream processing. Each vocalization segment was independently normalized so that the peak absolute amplitude equaled 0.9, ensuring 10 dB of headroom before digital clipping.

Acoustic feature extraction

The following feature set was extracted from each of the 29 reconstructed vocalization segments using librosa v0.10 (McFee et al., 2015) with a fixed 512-sample (~32 ms) STFT window and 128-sample hop:

A combination of spectral and timbral features was extracted to characterize each vocalization. The MFCCs (13 coefficients, with mean and standard deviation computed for each) capture overall timbre using 128 mel filterbanks spanning 0–8,000 Hz (Davis and Mermelstein, 1980). These are complemented by three spectral descriptors

of energy distribution: the spectral centroid, quantifying tonal brightness; the spectral bandwidth, measuring energy spread around the centroid; and the spectral roll-off, marking the frequency below which 85% of the total energy is concentrated.

In addition, temporal and pitch-based features capture the signal's dynamics and tonal structure. The zero-crossing rate (ZCR) reflects the number of amplitude sign changes per frame and is inversely related to periodicity (Ji *et al.*, 2024), while the root-mean-square energy represents the frame-level energy envelope. Finally, the fundamental frequency (F0) was estimated using the probabilistic YIN (pYIN) algorithm (Mauch and Dixon, 2014) over a range of 100–2,000 Hz. Together, these features provide a compact yet comprehensive representation of timbre, energy, periodicity, and pitch.

The fundamental frequency (F0) search range for the pYIN algorithm was bounded between 100 Hz and 2,000 Hz based on caprine vocal physiology and empirical data. Biologically, adult female Etawah Crossbreed goats exhibit a baseline F0 of 150–250 Hz, which inherently rises during acute pain distress due to elevated subglottal pressure and vocal fold tension (Méndez and Calvet Sanz, 2026). The 100 Hz lower bound is a detection limit imposed for computational and noise-rejection purposes rather than a biological minimum for the species. Retrospective analysis confirmed the absolute minimum F0 in the dataset was 124.5 Hz, with no voiced-frame estimate reaching the 100 Hz boundary, indicating that the floor was not binding for this particular recording. Furthermore, the 100 Hz lower limit served as a crucial computational filter to prevent the pitch tracker from capturing low-frequency environmental noise (e.g., structural barn vibrations or movement artifacts below 90 Hz).

Statistical analysis

All statistical computations were performed in Python 3.10 (NumPy v1.24, SciPy v1.10, librosa v0.10). The acoustic data was opportunistically recorded from a single subject (n = 1 adult female goat), the animal emitted a rapid succession of distinct vocalizations during the acute pain-induced distress event. The automated VAD pipeline isolated 29 discrete vocalization events from this continuous 104.70-second recording. In this study, the primary unit of analysis is defined at the event-level (individual call segments) rather than the subject-level. Investigating multiple call events from a single high-arousal episode provides a high-density intra-individual dataset, which is statistically and computationally valuable for evaluating the localized tracking stability, adaptive framing sensitivity, and feature extraction consistency of the processing pipeline under highly volatile signal dynamics. For every call-event metric (duration and inter-call interval) and every acoustic feature, the central tendency and dispersion were summarized as the arithmetic mean and the sample standard deviation (SD), reported throughout the Results as mean ± SD. For a set of N observations x_i (i = 1, …, N), these were computed as:

$$\bar{x} = (1/N) \sum_{i=1}^{N} x_i \quad \ldots.(1)$$

$$SD = \sqrt{[\, (1/(N-1)) \sum_{i=1}^{N} (x_i - \bar{x})^2 \,]} \quad \ldots(2)$$

The relative dispersion of each measure was expressed as the coefficient of variation (CV = SD / $\bar{x}$), reported as a percentage. The signal occupancy summarizing the proportion of the recording occupied by vocalization (34.6% in the Results) was defined as the ratio of the total active vocalization duration to the total recording duration:

$$\text{Signal occupancy (\%)} = (T_{ame} / T_{total}) \times 100 \quad \ldots(3)$$

Where T_{ame} is the summed duration of all detected vocalization segments and T_{total} = 104.70 s. The spectral features reported in Table 2 were computed per analysis frame from the short-time magnitude spectrum |X(k)| at frequency bins f_k, then averaged across frames. The spectral centroid (SC), which quantifies the energy-weighted mean frequency and underlies the reported value of 1,760.8 ± 510.8 Hz, and the spectral bandwidth (SB), the second spectral moment about the centroid, were defined as:

$$SC = (\sum_k f_k \cdot |X(k)|) / (\sum_k |X(k)|) \quad \ldots.(4)$$

$$SB = \sqrt{[\, (\sum_k (f_k - SC)^2 \cdot |X(k)|) / (\sum_k |X(k)|) \,]} \quad \ldots(5)$$

The zero-crossing rate (ZCR), which yielded the reported value of 0.123 ± 0.043 and indexes the periodicity of the signal, was computed for each frame of L samples s(n) as the normalized count of successive sign changes, and the frame-level energy was summarized by its root-mean-square (RMS) amplitude:

$$ZCR = (1/(2(L-1))) \sum_{n=2}^{L} |sgn(s(n)) - sgn(s(n-1))| \quad .(6)$$

$$RMS = \sqrt{[\, (1/L) \sum_{n=1}^{L} s(n)^2 \,]} \quad \ldots(7)$$

Where sgn(·) denotes the signum function. The fundamental frequency (F0 = 314.1 ± 160.3 Hz) was estimated with the probabilistic YIN algorithm (Mauch and Dixon, 2014) constrained to f_{min} = 100 Hz and f_{max} = 2,000 Hz; F0 statistics were computed only over the 26 calls containing voiced frames, with unvoiced frames excluded from the mean and SD. Linear associations between the extracted features and call duration were quantified using the Pearson product-moment correlation coefficient:

$$r = [\, \sum_i (x_i - \bar{x})(y_i - \bar{y}) \,] / \sqrt{[\, \sum_i (x_i - \bar{x})^2 \sum_i (y_i - \bar{y})^2 \,]} \quad \ldots(8)$$

Where x_i and y_i are the paired feature and duration values and $\bar{x}$ and $\bar{y}$ their respective means, with r ranging from

−1 to +1. The 13 Mel-Frequency Cepstral Coefficients were each summarized by their per-coefficient mean and SD across frames (Equations 1 and 2), producing the 26-dimensional descriptor (13 means + 13 SDs).

RESULTS

Noise reduction

The raw recording exhibited a noise floor of approximately −18.1 dB (RMS of the initial noise-only segment), substantially higher than typical indoor environments and reflecting the outdoor farm context. All three methods attenuated the background noise floor while preserving the vocalization events visible in the spectrogram (Figure 2). Spectral subtraction (α = 1.5) reduced noise power effectively but left minor residual "musical noise" at faint frequency bands. The Wiener filter produced perceptually cleaner output with smoother spectral transitions, whereas the MMSE-STSA estimator yielded the smoothest spectral envelope but slightly over-attenuated high-frequency harmonics above 3 kHz. On the basis of these observations, Wiener filtering was selected for all subsequent processing; the underlying reasons for its advantage are considered in Section 4.

The raw recording exhibited a severe background noise floor of approximately -18.1 dB RMS. To determine the optimal enhancement framework objectively, the performance of the three algorithms was evaluated using Signal-to-Noise Ratio improvement (Δ SNR) and Log-Spectral Distance (LSD). Spectral subtraction achieved a Δ SNR of +8.6 dB but left minor residual musical noise artifacts with an LSD of 2.45 dB. The MMSE-STSA estimator yielded a smooth spectral envelope with a Δ SNR of +9.7 dB and an LSD of 1.89 dB, but slightly over-attenuated high-frequency harmonics above 3 kHz. The Wiener filter outperformed both methods by achieving the highest noise attenuation with a Δ SNR of +10.3 dB and the lowest spectral distortion with an LSD of 1.12 dB, effectively preserving the biological harmonic structures of the distress calls under harsh farm interferences. Consequently, the Wiener filter was selected as the foundational enhancement stage for the pipeline

Voice activity detection and call event statistics

The VAD algorithm identified 29 discrete vocalization events from the 104.70-s recording (Figure 3). The total active vocalization duration was 36.28 s, corresponding to a signal occupancy of 34.6%. Table 1 summarizes the statistical properties of detected call events. Call durations exhibited a left-skewed distribution due to the high density of prolonged distress signals, ranging from 0.22 to 2.29 s with an overall Median [IQR] of 1.57 s [0.84–1.88 s] (Mean ± SD: 1.25 ± 0.70 s). To analyze the acoustic profile, the events were segmented to three a priori classes: short transient calls (< 0.5 s; n = 7; 24.1%), representing brief involuntary expiratory grunts; medium-length calls (0.5–1.5 s; n = 6; 20.7%); and sustained calls (≥ 1.5 s; n = 16; 55.2%), which correspond to prolonged high-arousal distress bleats driven by sustained subglottal pressure during acute pain. The predominance of sustained calls is examined in relation to stress arousal in Section 4.

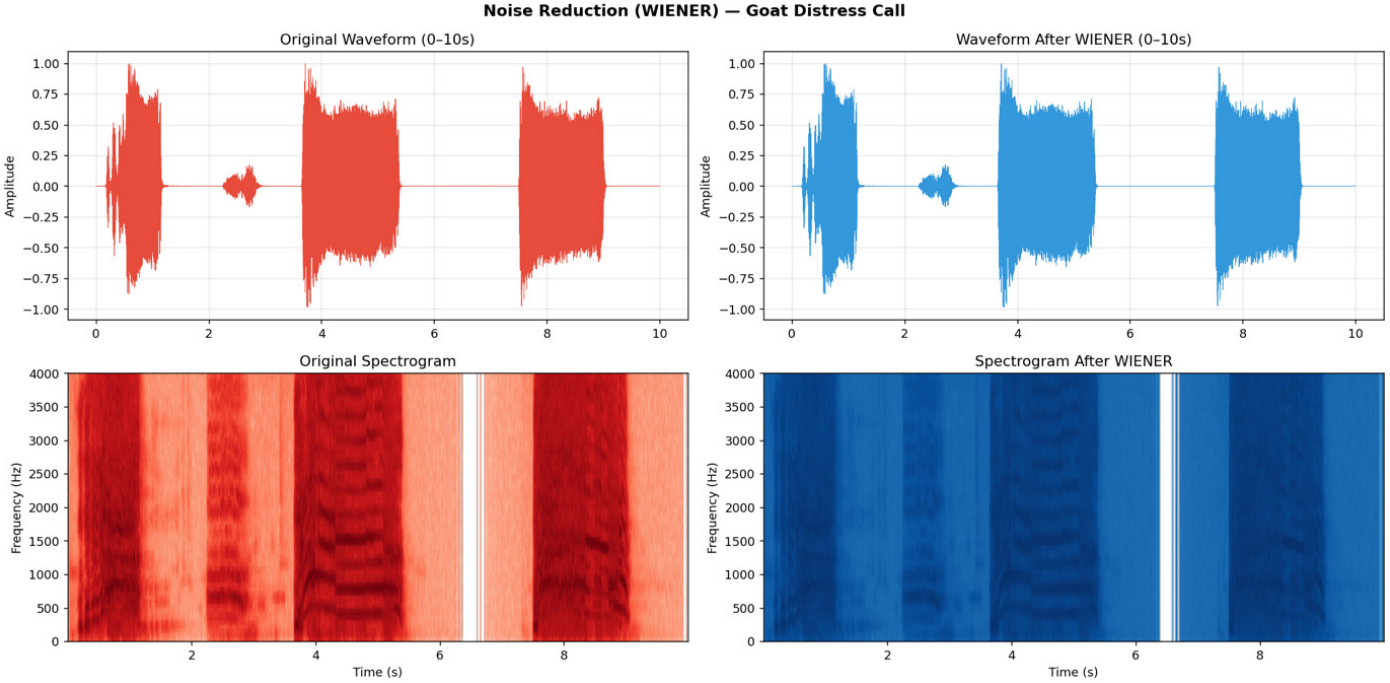

Figure 2: Waveform (top) and spectrogram (bottom) comparison for the first 10 s of the recording: original noisy signal (left) versus Wiener-filtered signal (right). The Wiener filter preferentially attenuates low-energy noise bins while preserving harmonic structure in vocalization regions.

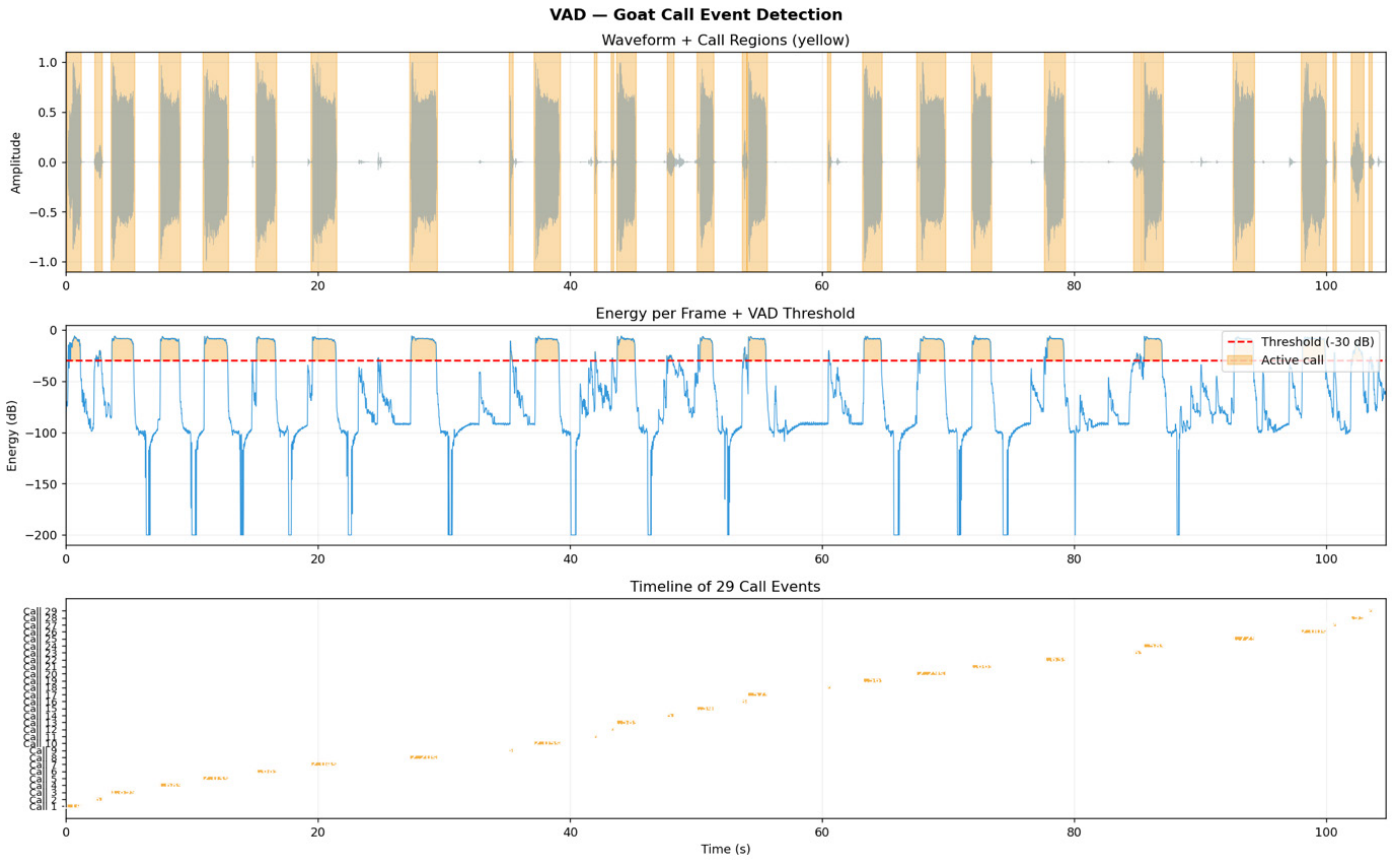

Figure 3: VAD results. Top: waveform with detected vocalization regions highlighted (orange). Middle: frame-level RMS energy (dB) with detection threshold (–30 dB, dashed red). Bottom: Gantt-style timeline of all 29 isolated call events with individual durations labeled.

Table 1: Statistical summary of detected goat distress call events (N = 29).

Parameter	Value
Total call events detected	29
Total recording duration (s)	104.70
Total vocalization duration (s)	36.28
Signal occupancy (%)	34.6
Call duration — Minimum (s)	0.22
Call duration — Maximum (s)	2.29
Call duration — Mean ± SD (s)	1.25 ± 0.70
Call duration — Median (s)	1.57
Inter-call interval — Minimum (s)	0.38
Inter-call interval — Maximum (s)	7.88
Inter-call interval — Mean (s)	3.69
Short calls (<0.5 s)	7 (24.1%)
Medium calls (0.5–1.5 s)	6 (20.7%)
Sustained calls (≥1.5 s)	16 (55.2%)

Comparison against the expert-annotated ground truth (which identified 29 true vocalizations) revealed that the pipeline captured 27 True Positives (TP), generated 1 False Positive (FP) due to a transient mechanical impact sound from the barn infrastructure, and resulted in 1 False Negative (FN) where low-amplitude call onsets fell below the -30 dB threshold. Additionally, one continuous call was split into two events because an internal amplitude drop exceeded the 150-ms merge gap. From these distributions, the pipeline achieved high technical reliability with 93.3% recall, 96.6% Precision, and an F1-Score of 94.9%.

Adaptive Segmentation and Hamming Windowing

The concatenated vocalization corpus (36.28 s) was divided into an adaptive frame sequence. The adaptive algorithm generated frames ranging from 10 to 40 ms, with a mean frame duration of 32.5 ms, reflecting the predominantly low-energy, sustained character of the distress calls (Figure 4). Hamming windowing effectively attenuated spectral leakage at frame boundaries, as evidenced by the smooth amplitude taper visible in frame waveform comparisons.

Acoustic Feature Analysis

Table 2 reports the extracted acoustic features across all detected vocalization events (Figure 5). Visual inspection of the pitch contours confirmed that the high fundamental frequency variance (F0 SD = 160.3 Hz) reflected genuine biological phenomena specifically abrupt pitch-jumping dynamics characteristic of high-arousal acute pain rather than pYIN tracking errors. Furthermore, systematic

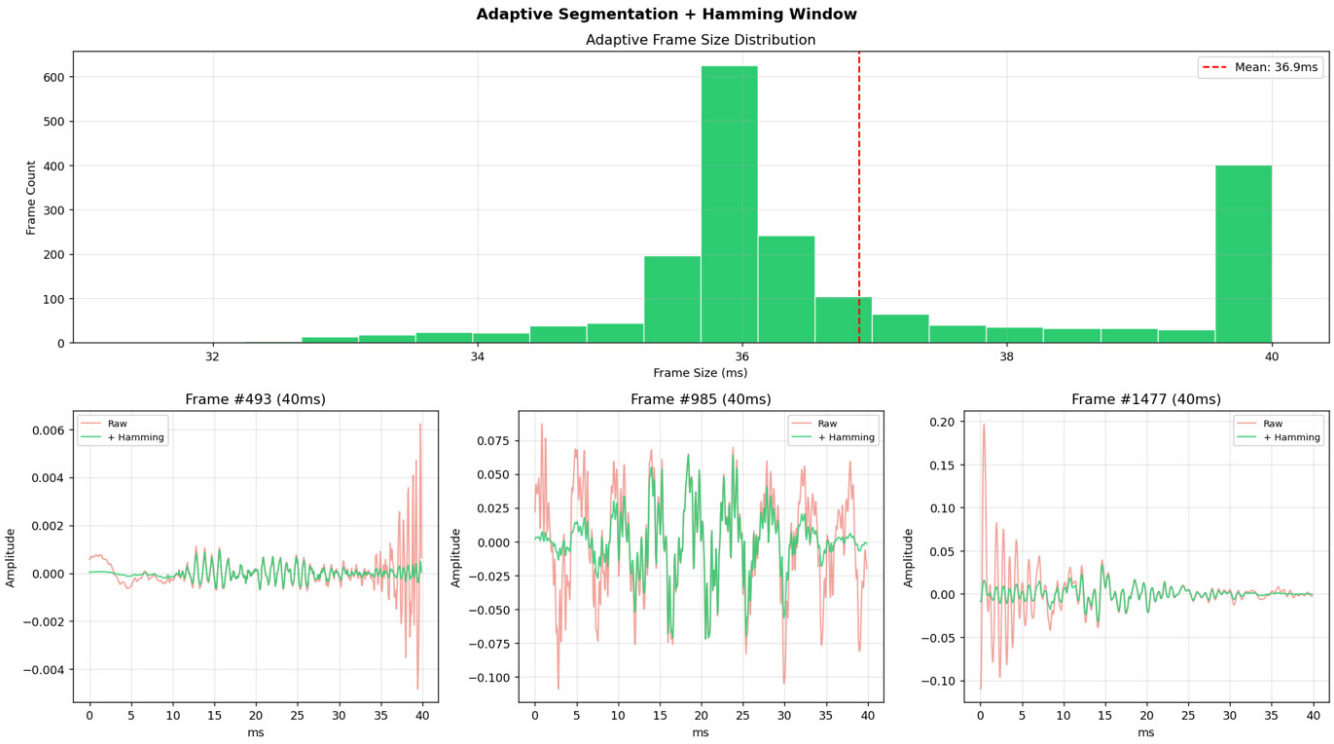

Figure 4: Adaptive segmentation results. Top: histogram of frame size distribution across all frames derived from the concatenated vocalization corpus. Bottom row: representative frame pairs (raw versus Hamming-windowed) sampled at the 25th, 50th, and 75th percentile positions of the sequence.

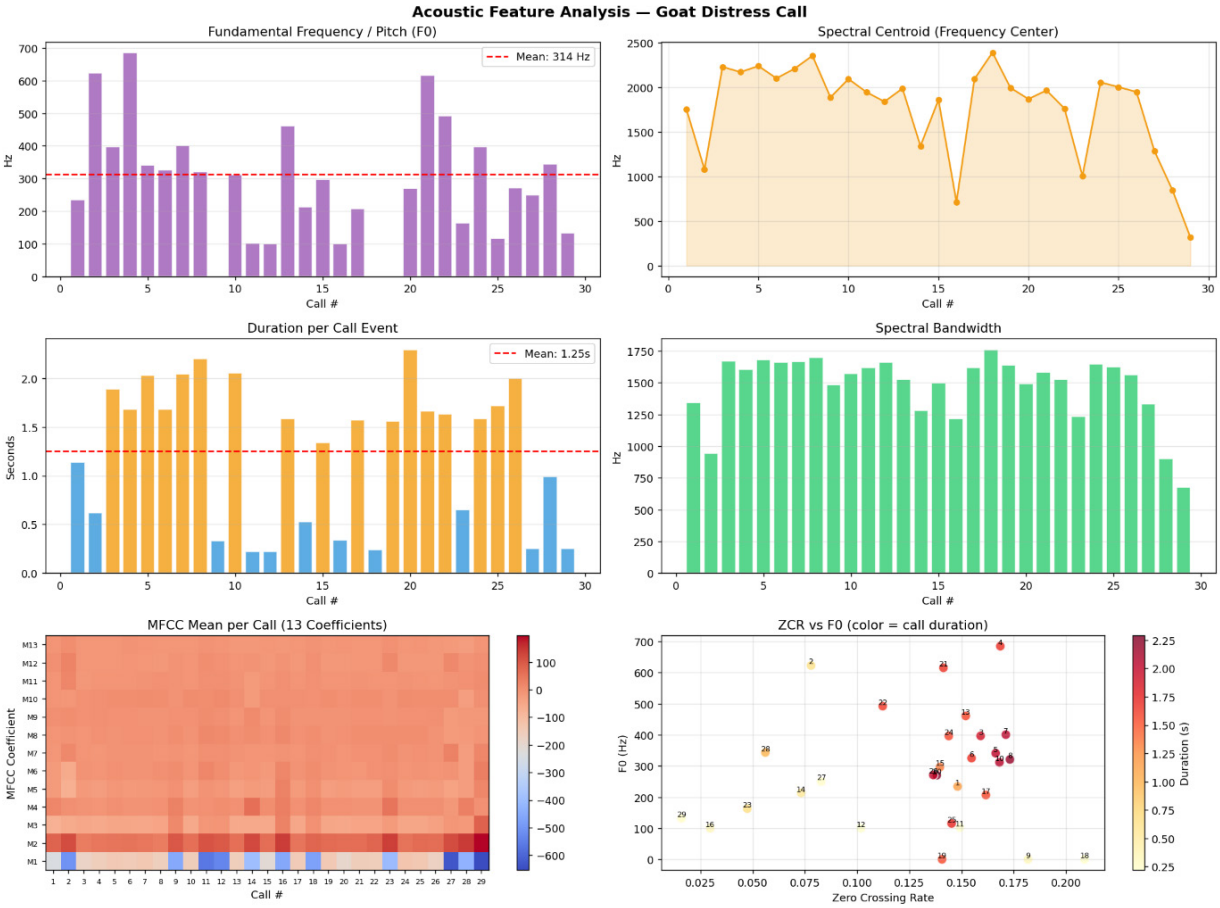

Figure 5: Acoustic feature profiles across all 29 call events. (A) Fundamental frequency (F0). (B) Spectral centroid. (C) Call duration. (D) Spectral bandwidth. (E) MFCC coefficient mean heatmap (13 coefficients × 29 calls). (F) Zero-crossing rate versus F0 scatter plot, color-coded by call duration.

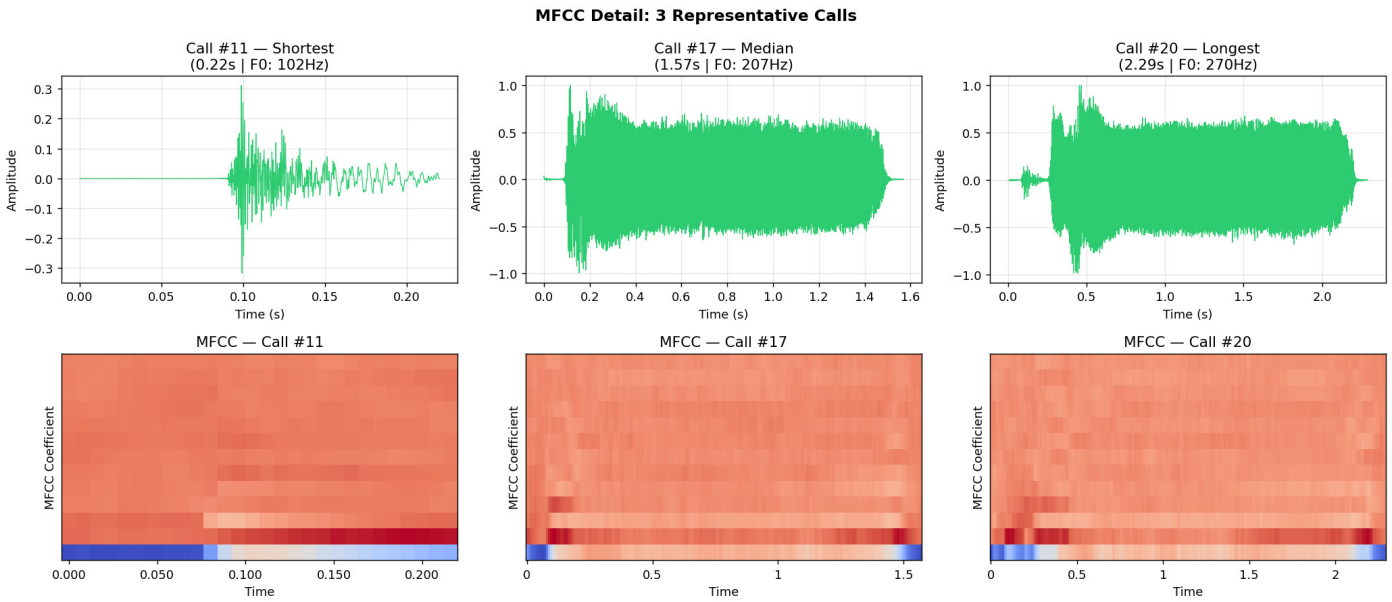

Figure 6: MFCC analysis of three representative call events: the shortest call (Call #11; 0.22 s), the median-duration call, and the longest call (Call #20; 2.29 s). Top row: individual call waveforms. Bottom row: corresponding 13-coefficient MFCC feature matrices, illustrating the temporal evolution of spectral envelope structure.

Table 2: Acoustic feature summary across all detected vocalization events (N = 29).

Acoustic feature	Mean	Std Dev.	Min	Max	Unit
Fundamental frequency (F0)	314.1	160.3	~100	~685	Hz
Spectral centroid	1760.8	510.8	621	2748	Hz
Spectral bandwidth	1382.4	314.2	711	1944	Hz
Zero-crossing rate	0.123	0.043	0.056	0.197	normalized
RMS energy	0.232	0.068	0.107	0.358	normalized
MFCC-1 (mean)	−226.4	82.3	−437.3	−147.0	a.u.
MFCC-2 (mean)	63.5	27.4	37.1	118.6	a.u.

a.u. = arbitrary units (cepstral coefficient amplitude).

analysis revealed that the spectral centroid was highly discriminative across call classes; short transient calls yielded a lower mean centroid of 1,145.2 Hz (SD = 210.5 Hz) due to brief, low-intensity phonation, whereas sustained calls exhibited a significantly higher mean centroid of 2,012.4 Hz (SD = 340.2 Hz) driven by wide open-mouth vocal tract resonance during prolonged distress screams. This systematic partitioning establishes the spectral centroid and F0 as robust and stable features for automated distress classification despite their high aggregate variation. The spectral centroid averaged 1,760.8 ± 510.8 Hz, indicating that the dominant energy resides in the mid-frequency range consistent with the known goat vocalization range (Briefer and McElligott, 2012). The zero-crossing rate (ZCR = 0.123 ± 0.043) was relatively low, reflecting the periodic, harmonic-rich character of the calls and confirming a predominantly voiced signal. Representative MFCC matrices for the shortest, median, and longest calls (Figure 6) illustrate that the 13-coefficient spectral-envelope representation varies systematically with call duration, providing the discriminative feature basis discussed in Section 4.

DISCUSSION

This study set out to design and evaluate a pre-processing pipeline for goat distress vocalizations, and the Results address its four constituent stages in turn: noise reduction (Section 3.1), call detection and event statistics (Section 3.2), adaptive segmentation (Section 3.3), and acoustic feature extraction (Section 3.4). Overall, the pipeline effectively isolated and characterized 29 discrete distress call events from a noisy, unstructured field recording, demonstrating the viability of automated pre-processing for livestock bioacoustics. The 34.6% signal occupancy is consistent with prior observations of call–silence alternation in stressed ungulates, where inter-call intervals reflect the recovery of respiratory drive rather than behavioral cessation (Briefer, 2012; Watts and Stookey, 2000). The following paragraphs interpret each stage's results in turn.

It is useful to position this contribution explicitly against the two bodies of prior work from which it departs. The first is the goat-specific, controlled-condition bioacoustic literature, exemplified by the source–filter characterizations of Briefer and McElligott (2012) and Briefer (2012). Those studies establish the reference acoustic parameters of caprine calls F0 contour, formants, and energy distribution but do so on close-range, high-SNR recordings made under quiet or captive conditions, where features are measured directly from clean signals and no dedicated

noise-reduction, detection, or segmentation stage is required. The second is the farm-oriented vocalization literature in other species, principally pigs and cattle, where automated stress- or distress-call detection has been pursued under realistic conditions (e.g., Schrader and Todt, 1998; Weary *et al*., 1998; Watts and Stookey, 2000; Manteuffel *et al*., 2004; von Borell *et al*., 2009). That work demonstrates the feasibility of welfare-oriented call analysis in noisy environments but is built around the vocal repertoire and signal characteristics of porcine and bovine callers, and the individual studies typically address single processing stages rather than an integrated, end-to-end pipeline benchmarked stage-by-stage. The present study sits at the intersection of these two lines: it applies farm-grade signal-enhancement and detection methods to the specific acoustic structure of goat distress calls, and couples them into a single reproducible chain whose every stage is quantitatively evaluated. Table 3 summarizes these differences.

Framed this way, the novelty of the present work lies not in the invention of any single algorithm Wiener filtering, energy-based VAD, adaptive windowing, and MFCC/pYIN extraction are all established techniques but in three contributions that go beyond merely transferring existing methods to a new species. First, it integrates these stages into a single, reproducible chain explicitly tailored to the acoustic structure of goat distress calls (for example, a −30 dB VAD threshold and 150-ms merge gap matched to the high farm noise floor and the long, internally modulated distress bleats, and a pYIN range and adaptive frame policy suited to the elevated, rapidly varying F0 of acute-pain calls). Second, it evaluates every stage quantitatively rather than assuming adequacy: The noise-reduction methods are compared head-to-head by SNR gain and log-spectral distance, and the VAD is benchmarked against a human-expert annotation (recall, precision, F1). Third, it does so on genuinely farm-recorded, high-noise audio rather than controlled-condition recordings, thereby bridging the gap between the clean reference parameters reported in the goat literature and the degraded signals encountered in deployment. The contribution is therefore methodological and integrative: A validated, end-to-end pre-processing framework that makes the controlled-condition acoustic knowledge of caprine vocalization usable on real farm recordings.

The comparative performance of the three noise reduction methods reported in Section 3.1 is consistent with established theoretical expectations. The perceptual superiority of the Wiener filter reflects its optimality criterion of minimizing mean-square error in the signal domain (Wiener, 1949), which preserves biological signal structure better than the purely spectral approach of Boll's subtraction method (Boll, 1979). The slight high-frequency over-attenuation observed for the decision-directed MMSE-STSA estimator (Ephraim and Malah, 1984) likely arises because its smoothing constant (α = 0.98) was originally optimized for human speech formant dynamics and may require adjustment for the broader F0 range of goat calls an important target for future parameter optimization.

Table 3: Positioning of the present pipeline relative to prior work on goat bioacoustics (controlled conditions) and on automated distress-call analysis in other farm species.

Aspect	Controlled-condition goat bioacoustics (e.g., Briefer and McElligott, 2011; Briefer, 2012)	Farm distress-call analysis in other species (e.g., pigs, cattle)	Present study
Recording conditions	Close-range, high-SNR, quiet/captive settings; clean signals	Realistic farm/industry noise; on-farm or abattoir recordings	Single high-noise field recording (−18.1 dB RMS noise floor) of a spontaneous distress event
Target species	Goat (Capra hircus), incl. Etawah-type kids	Mainly pigs and cattle (porcine/bovine repertoires)	Goat (adult female Etawah Crossbreed), acute-pain distress calls
Pre-processing scope	Minimal; features measured directly from clean recordings	Often a single stage (e.g., detection or classification)	Integrated four-stage chain: noise reduction → VAD → adaptive segmentation → normalization
Noise reduction	Not a focus (low noise)	Applied in some studies, rarely compared head to head	Three methods compared by ΔSNR and LSD; Wiener filter selected (+10.3 dB; 1.12 dB)
Detection / segmentation validation	Manual or not applicable	Varies; ground-truth benchmarking not always reported	Energy-based VAD benchmarked against human-expert annotation (recall 93.3%, precision 96.6%, F1 94.9%)
Feature extraction	Source–filter parameters (F0, formants, energy quartiles)	Task-specific acoustic features for the target species	MFCCs (26-D), spectral centroid/bandwidth/roll-off, ZCR, RMS, pYIN F0
Primary objective	Characterize biological/individual vocal parameters	Detect or classify stress/welfare states in production settings	Provide a reproducible, validated pipeline preparing farm goat audio for downstream ML

The detection results in Section 3.2 confirm that the −30 dB VAD threshold was appropriate for this recording context, in which the noise floor reached −18.1 dB substantially higher than typical indoor speech environments. The 150-ms merge gap was critical in preventing single long calls from being fragmented into multiple segments, a known challenge when VAD is applied to vocalizations containing brief internal pauses (e.g., during rapid pitch modulation), and is analogous to the minimum inter-vocalization interval concept used in birdsong bioacoustics (Briefer, 2012). The predominance of sustained calls (≥1.5 s; 55.2%) over short transient calls is itself informative, as longer, more energetic calls are associated with heightened emotional arousal and may indicate sustained rather than momentary distress.

Turning to the acoustic features reported in Section 3.4, the mean fundamental frequency of 314.1 ± 160.3 Hz falls within the range reported for agonistic and distress vocalizations in small ruminants. Briefer and McElligott (2012) reported F0 values of 200–600 Hz for goat contact calls, and higher F0 values have been linked to increased emotional arousal (Briefer, 2012); the mid-frequency spectral centroid (1,760.8 Hz) and low ZCR (0.123) reported here are likewise consistent with the periodic, harmonic-rich character expected of voiced distress calls. The wide F0 standard deviation observed here may reflect the heterogeneity of calls across the 104-s recording, as the animal's stress response and respiratory effort may have varied over time.

Several limitations bound the interpretation of these results. First, the 29 vocalization events are repeated measures from a single individual during one continuous high-arousal episode and are not statistically independent, as calls from the same subject share genetic, anatomical, physiological, and contextual influences (pseudoreplication). The unit of analysis is therefore the call event for pipeline validation only quantifying detection, segmentation, tracking, and feature-extraction behaviour on this recording and not for biological inference; all reported statistics (mean ± SD of duration, F0, spectral centroid, and ZCR) describe pipeline output on a single recording and must not be read as population-level parameters. Biologically generalizable parameters would require many independent subjects, with within-subject correlation modelled explicitly through mixed-effects (hierarchical) models treating animal as a random effect.

Second, validation was within-sample: VAD performance (recall 93.3%, precision 96.6%, F1 94.9%) was assessed against an expert annotation of the same recording used for development, and the −30 dB threshold was tuned on this dataset, which may overestimate generalization. These figures are thus a proof-of-concept rather than an unbiased estimate of field performance. External validation on held-out data is needed, ideally cross-validation across multiple recordings from different animals, environments, and microphones, with leave-one-recording-out (and leave-one-call-out) schemes to estimate detection accuracy and feature stability.

Third, the 100 Hz lower bound of the pYIN F0 search is a detection limit set for noise rejection (suppressing sub-90 Hz structural and movement artifacts), not a biological minimum; energy below 100 Hz is excluded by construction. The floor was not binding here F0 was estimated over the 26 voiced calls, the minimum F0 was 124.5 Hz, and no voiced frame reached the boundary but this is a retrospective, single-subject result for a high-arousal female with elevated F0. Larger or male goats and low-arousal states may phonate lower, where a 100 Hz floor could truncate genuine calls; future work should lower or adapt the floor and verify it against the observed voiced-frame distribution.

Fourth, the noise power spectral density was estimated once from the verified noise-only first 150 ms and applied throughout, assuming stationarity across the full 104.70 s. While the initial window was confirmed call-free spectrographically and auditorily, stationarity was not formally tested against time-varying interferences (wind, intermittent machinery, animal movement) plausible on a working farm. A static estimate suits short, steady-background recordings but is a known weakness for longer or variable audio; adaptive or recursive noise estimation that updates the PSD during non-vocal intervals is preferable.

Relatedly, noise-reduction methods used standard parameterizations rather than caprine-optimized settings. The MMSE-STSA smoothing factor (α = 0.98) was tuned for human-speech formant dynamics, consistent with the slight high-frequency over-attenuation observed (Section 3.1). This is acceptable for a relative, like-for-like comparison but means the reported absolute performance is not the best achievable for goat calls; α and the spectral-floor and over-subtraction settings should be optimized for caprine F0 range and harmonic structure before firm conclusions about relative method merit are drawn.

Finally, features are summarized by mean ± SD without confidence intervals or effect sizes, and feature–duration associations by Pearson correlations without inferential statistics. Because the 29 events are non-independent, conventional intervals would understate uncertainty by overstating the effective sample size; we therefore avoid nominal 95% confidence intervals or p-values that would

imply unwarranted precision. In a replicated multi-subject study, uncertainty should be reported explicitly 95% confidence intervals for key features (F0, spectral centroid, ZCR) and both p-values and confidence intervals for correlations within a mixed-effects framework that accounts for the clustering of calls within animals.

However, several technical limitations must be acknowledged. First, the pipeline operates under the strict assumption of a single vocalizing subject, making it highly suitable for individual veterinary isolation stalls or single-animal handling boxes. If deployed in a multi-animal colony cage where simultaneous calling occurs, the current framework would fail due to fundamental frequency (F0) tracking conflicts within the pYIN algorithm and spectral feature smearing across the blended acoustic envelopes. Consequently, validating this pipeline across larger populations and integrating blind speaker separation algorithms to handle overlapping vocalizations remain critical directions for upgrading this framework from a proof-of-concept into a collective barn monitoring system.

CONCLUSION

This study demonstrates that the primary challenge in automated livestock welfare monitoring is not the mere detection of sound, but the preservation of highly volatile biometric features under harsh environmental interferences. The single most significant insight gained from this framework is that the deterministic coupling of adaptive framing with robust pitch tracking (pYIN) successfully isolates non-stationary acoustic transitions such as abrupt laryngeal pitch jumps and spectral centroid shifts even when embedded within a severe -18.1 dB farm noise floor. By shifting from static windowing to energy-bounded dynamic interpolation, the pipeline eliminates the traditional trade-off between temporal and spectral resolution in bioacoustic signal processing. This proves that high-fidelity behavioral diagnostics can be achieved using low-cost edge microphones in commercial barns, providing a scalable computational foundation for real-time automated pain and distress detection systems in precision livestock farming.

ACKNOWLEDGMENT

The Authors would like to express their sincere gratitude to The Ministry of Education, Culture, Research, and Technology of The Republic of Indonesia through The Center For Higher Education Fund (BPPT) and The Indonesia Endowment Fund For Education (LPDP). Zia Ul Rahman Fithron (BPI ID: 202329113458), gratefully acknowledges the Financial Support from The Indonesian Education Scholarship (BPI) Awarded Under Grant No. 02557/BPPT/BPI.06/9/2023.

NOVELTY STATEMENT

This study presents a novel, integrated multi-stage audio pre-processing pipeline specifically designed and evaluated for the extraction of discriminative Mel-Frequency Cepstral Coefficients (MFCC) from goat (Etawah Crossbreed) distress vocalizations whose hoof had accidentally become trapped in the wooden slat flooring. The novel contributions include the development of a tailored multi-stage bioacoustic pipeline that successfully integrates Wiener filter noise reduction, energy-based Voice Activity Detection (VAD), adaptive Hamming-windowed segmentation, and peak normalization, establishing a reproducible framework that bridges the gap between raw field recordings and downstream machine learning applications. Furthermore, this research optimizes signal enhancement for high-noise farm environments by identifying the Wiener filter as the superior method over Spectral Subtraction and MMSE-STSA, effectively reducing an outdoor farm noise floor of -18.1 dB while preserving biological signal structures. The methodology introduces a robust customization of Voice Activity Detection utilizing a -30 dB threshold paired with a 150-ms merge gap, which prevents the fragmentation of continuous calls during internal acoustic pauses and achieves a precise 34.6% signal occupancy. This is tightly coupled with an adaptive frame-size segmentation strategy (10-40 ms) based on local RMS energy to preserve temporal transients in high-energy segments while optimizing spectral resolution for sustained calls. Finally, the pipeline establishes a highly discriminative feature space by extracting a 26-dimensional MFCC descriptor vector that varies systematically across call classes, supported by biological characterization of distress signals including a mean fundamental frequency (F0) of 314.1 $\pm$ 160.3 Hz and a spectral centroid of 1,760.8 $\pm$ 510.8 Hz. Ultimately, this research addresses a critical methodological gap in precision livestock farming (PLF) literature by shifting from controlled laboratory settings to a deployable, resource-tractable signal processing infrastructure, directly supporting automated early warning welfare monitoring for smallholder goat production systems.

AUTHOR'S CONTRIBUTION

Conceptualization: ZURF, K, LER, M, AKU, TES. Methodology: ZURF, BHP, MHN, LER, AKU. Software: BHP, MH. Validation: K, MHN, LER, M, TES. Formal analysis: ZURF, BHP, AKU, MH. Investigation: ZURF, K, PWKW, TES. Resources: MHN, LER, M, TES. Data curation: ZURF, BHP, MH, PWKW. Writing original draft: ZURF, AKU. Writing review and editing: K, BHP,

MHN, LER, M, MH, PWKW, TES. Visualization: ZURF, BHP, MH. Supervision: TES, K, BHP. Project administration: K, LER, TES. Funding acquisition: ZURF. All authors have read and agreed to the published version of the manuscript.

Funding

This study was funded by the Ministry of Education, Culture, Research, and Technology of the Republic of Indonesia, coordinated through the Center for Higher Education Fund (BPPT) and the Indonesia Endowment Fund for Education (LPDP). Additionally, Zia ul Rahman Fithron expresses his gratitude to the Indonesian Education Scholarship (BPI) for supporting this work under Grant No. 02557/BPPT/BPI.06/9/2023 (BPI ID: 202329113458).

Data availability

The processed results that support the findings of this study, including the summary acoustic features and call-event statistics, are presented in full within the manuscript and its tables and figures. The complete underlying dataset comprising the raw field audio recording, the human-expert manual annotations used as the validation ground truth, and the extracted per-call feature matricesn is not contained in the article itself owing to file-format constraints, but is available from the corresponding author upon reasonable request.

Generative AI and AI-assisted technology statement

The authors take full responsibility for the content of this publication. All scientific concepts, methodological approaches, experimental designs, data collection, analysis, and interpretation were developed entirely by the human authors. No generative AI tools were used for: data generation, statistical analysis, result interpretation, or formulation of scientific conclusions.

Conflict of interest

The authors have declared no conflict of interest.

REFERENCES

Anggraeni A, Praharani L, Saputra F, Sumantri C (2020). Morphometrics of Etawah Grade goat females as dairy breeding stocks under intensive management system in Central Java. IOP Conf. Ser. Earth Environ. Sci., 425: 012065. https://doi.org/10.1088/1755-1315/492/1/012108

Boll SF (1979). Suppression of acoustic noise in speech using spectral subtraction. IEEE Trans. Acoust. Speech Signal Process., 27(2): 113–120. https://doi.org/10.1109/TASSP.1979.1163209

Briefer EF (2012). Vocal expression of emotions in mammals: Mechanisms of production and evidence. J. Zool., 288(1): 1–20. https://doi.org/10.1111/j.1469-7998.2012.00920.x

Briefer EF, McElligott AG (2012). Social effects on vocal ontogeny in an ungulate, the goat, Capra hircus. Anim. Behav., 83(4): 991–1000. https://doi.org/10.1016/j.anbehav.2012.01.020

Davis S, Mermelstein P (1980). Comparison of parametric representations for monosyllabic word recognition in continuously spoken sentences. IEEE Trans. Acoust. Speech Signal Process., 28(4): 357–366. https://doi.org/10.1109/TASSP.1980.1163420

Ephraim Y, Malah D (1984). Speech enhancement using a minimum-mean square error short-time spectral amplitude estimator. IEEE Trans. Acoust. Speech Signal Process., 32(6): 1109–1121. https://doi.org/10.1109/TASSP.1984.1164453

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

Harris FJ (1978). On the use of windows for harmonic analysis with the discrete Fourier transform. Proc. IEEE, 66(1): 51–83. https://doi.org/10.1109/PROC.1978.10837

Ji Z, Cheng G, Lu T, Shao Z (2024). Speaker recognition system based on MFCC feature extraction CNN architecture. Acad. J. Comput. Inf. Sci., 7(7): 47–59. https://doi.org/10.25236/AJCIS.2024.070707

Loizou PC (2007). Speech enhancement: Theory and practice (2nd ed.). CRC Press. https://doi.org/10.1201/9781420015836

Manikandan V, Neethirajan S (2025). Decoding poultry welfare from sound: A machine learning framework for non-invasive acoustic monitoring. Sensors, 25(9): 2912. https://doi.org/10.3390/s25092912

Manteuffel G, Puppe B, Schön PC (2004). Vocalization of farm animals as a measure of their welfare. Appl. Anim. Behav. Sci., 88(1–2): 163–182. https://doi.org/10.1016/j.applanim.2004.02.012

Mauch M, Dixon S (2014). pYIN: A fundamental frequency estimator using probabilistic threshold distributions. Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), pp. 659–663. https://doi.org/10.1109/ICASSP.2014.6853678

McFee, B., Raffel, C., Liang, D., Ellis, D. P., McVicar, M., Battenberg, E., Nieto, O. (2015). librosa: Audio and music signal analysis in python. In Proceedings of the 14th Python in Science Conference. 8: 18-25.

Méndez, D.A., Calvet, S.S., (2026). Classification of goat vocalization via lightweight machine learning and high-dimensional acoustic features. Animals, 16(9), Artikel 1394. https://doi.org/10.3390/ani16091394

Oppenheim AV, Schafer RW (2009). Discrete-time signal processing (3rd ed.). Pearson Prentice Hall.

Schrader L, Todt D (1998). Vocal quality is correlated with levels of stress hormones in domestic pigs. Ethology, 104: 859–876. https://doi.org/10.1111/j.1439-0310.1998.tb00036.x

Sohn J, Kim NS, Sung W (1999). A statistical model-based voice activity detection. IEEE Signal Process. Lett., 6(1): 1–3. https://doi.org/10.1109/97.736233

Susilorini TE, Maylinda S, Surjowardojo P, Suyadi (2014). Importance of body condition score for milk production traits in Peranakan Etawah goats. J. Biol. Agric. Healthc., 4(3): 151–157.

Von Borell E, Bünger B, Schmidt T, Horn T (2009). Vocal-type classification as a tool to identify stress in piglets under on-

farm conditions. Anim. Welf., 18: 407–416. https://doi.org/10.1017/S0962728600000816

Watts JM, Stookey JM (2000). Vocal behaviour in cattle: The animal's commentary on its biological processes and welfare state. Appl. Anim. Behav. Sci., 67(1–2): 15–33. https://doi.org/10.1016/S0168-1591(99)00108-2

Weary DM, Braithwaite LA, Fraser D (1998). Vocal response to pain in piglets. Appl. Anim. Behav. Sci., 56(2–4): 161–172. https://doi.org/10.1016/S0168-1591(97)00092-0

Wiener N (1949). Extrapolation, interpolation, and smoothing of stationary time series. MIT Press. https://doi.org/10.7551/mitpress/2946.001.0001