International Journal of Speech and Audiology

E-ISSN: 2710-3854
P-ISSN: 2710-3846
IJSA 2025; 6(2): 01-06

Journal's Website
Received: 10-04-2025
Accepted: 15-05-2025

Mahi Shah
Centennial High School (12th grade), Frisco, TX, USA

Shriya Nadagouda
Greenhill School (12th grade), Frisco, TX, USA

Aanya Naini
Hockaday School (12th grade) Dallas, TX 75229

Correspondence
Mahi Shah
Centennial High School (12th grade), Frisco, TX, USA

Acoustic classification of canine Vocalizations using spectrogram-based machine learning to infer emotional states

Mahi Shah, Shriya Nadagouda and Aanya Naini

DOI: https://www.doi.org/10.22271/27103846.2025.v6.i2a.78

Abstract
Dogs use barking as their primary form of communication, conveying emotions and needs much like humans use speech. However, misinterpretation of these vocalizations often results in harmful corrective measures, such as shock collars or devocalization surgeries. (Ehasni *et al*. 2018) This study developed a machine learning-based system that analyzes dog barks through spectrogram data and classifies them into emotional subtypes, including fear, aggression, and playfulness. A model trained using Apple's Create ML achieved an overall accuracy of 93.2%, with class-specific accuracies of 93% for fear, 95% for aggression, and 97% for playfulness. These results suggest that machine learning can provide a reliable and humane alternative for interpreting canine emotions, potentially reducing the use of harsh behavioral interventions.

Keywords: Canine vocalization, barking, spectrogram, machine learning, dog emotion detection, audio classification, Create ML, Acoustic

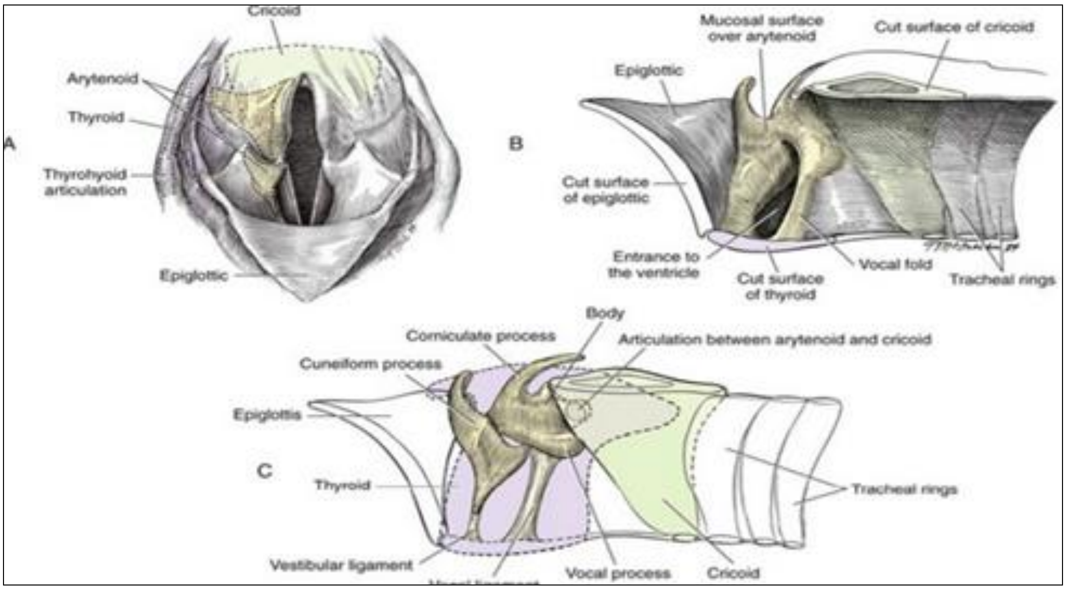
Introduction
Barking is a dog's main method of communication. However, its interpretation by humans is often flawed, leading to negative outcomes. Barking perceived as a nuisance can result in extreme actions such as shock collar usage or Ventriculocordectomy—surgical devocalization. To address this, we developed a humane, software-based tool to decode barking patterns and associate them with emotional states using spectrogram analysis and supervised machine learning.

Background
Limitations of Surgical and Aversive Approaches to Bark Suppression
Ventriculocordectomy is the veterinary medical term for the devocalization procedure. (Sophia 20400123X) When the surgery is performed for pet owner convenience, the goal is to eliminate dog barking through a surgical removal of the vocal cords. This surgery is not at all a minor surgery. It is an invasive procedure which has risks of blood loss, infection and other serious issues.
Because the focus of this surgical trauma is in the area through which the animal normally receives all air, fluids and food, the recovery process after ventriculocordectomy requires very close attention to hemorrhage, coughing and gagging. It has both long-term physical and psychological risks to the dog including increased levels of frustration and stress, compromised airway access, threats to physical safety because of the inability to ward off threats and dangers through vocalizing and engenders inappropriate behaviors.
A certified veterinarian writes, "Dogs bark; that's what they do. There is always a reason why they bark that should be understood and addressed. A surgical solution is not the answer and furthermore, it's inhumane."

Shock collars: Devices that deliver electric stimuli of varying intensities-are among the most used aversive tools to suppress undesirable canine behaviors, including excessive barking. These devices function by associating barking with physical discomfort, often through mild to painful shocks. While they may produce short-term behavioral suppression, they do not address the underlying cause of the vocalization and can lead to long-term behavioral and psychological consequences.

The use of shock collars, along with other aversive devices such as citronella and ultrasonic collars, has been widely criticized for its potential to induce fear, anxiety, and even aggression in dogs. These punishment-based methods disrupt the bond between humans and animals and may result in generalized behavioral problems. Misuse is prevalent, especially in untrained hands, leading to inconsistent correction, increased stress levels, and compromised welfare.

Modern animal behavior science increasingly supports reward-based training methods, emphasizing humane and effective behavior modification strategies. Numerous countries and regions have banned or strictly regulated the use of shock collars and similar aversive tools, recognizing that the harm they inflict outweighs any perceived benefits.



The larynx is known as the voice box, consisting of the vocal cords and cartilages that provide the semi-flexible structure of the larynx. Dogs have a modifiable vocal tract, meaning they have very plastic vocal cords. These vocal cords are composed of ligaments and muscles, covered with mucosal tissue. Dogs can slightly alter their voices to produce a wide variety of ranging sounds. These ranging barks have different meanings. The variations in bark structure can be observed through spectrogram analysis and conclude that these barks can be divided into contextual subtypes.

Methodology

To address the problem of misinterpreted and excessive dog barking-often leading to inhumane interventions such as shock collars or devocalization surgeries—we propose a humane, software-based solution grounded in machine learning. The core of our solution is an emotion classification model trained using Apple's Create ML, capable of decoding canine vocalizations based on acoustic features extracted from bark spectrograms. (Matsumi Suzuki G10L17/26)

This machine learning model is designed to recognize emotional states such as loneliness, fear, aggression, pain, playfulness, and alertness through supervised learning on annotated spectrogram data. By identifying the emotional context behind a dog's bark, this tool allows pet owners to respond more appropriately and compassionately, thereby reducing behavioral issues and enhancing animal welfare.

Our solution consists of the following key components

Data Collection: A total of 320 dog bark audio samples were compiled from multiple sources to ensure a diverse and representative dataset. Primary sources included subscription-based public databases such as the Dog Voice Emotion Dataset on Kaggle and professional-grade audio samples from Pond5. In addition, custom bark recordings were curated from diverse real-world environments using video capture tools. This approach allowed for the inclusion of a broader range of behavioral contexts and acoustic conditions, enhancing the ecological validity of the dataset.

Each audio sample was manually labeled according to its inferred emotional category—such as *fear*, *aggression*, *playfulness*, *pain*, *alert*, and *loneliness*—based on the situational context and behavioral cues observed during data capture. This emotion-specific labeling formed the ground truth for subsequent machine learning classification and ensured a robust foundation for supervised training.

Data Processing

The raw audio samples underwent a multi-step preprocessing pipeline to prepare them for spectral analysis and machine learning training. Initial noise reduction and audio segmentation were performed using WeVideo, a cloud-based video editing software. This tool was utilized to remove non-bark segments, such as ambient noise and prolonged silences, while preserving the natural intervals between bark bursts.

Post-cleaning, the audio data was converted into spectrograms using Audacity, a digital audio editing software. Spectrograms represent time-frequency distributions of sound signals and enable detailed visual and quantitative analysis of bark characteristics, including pitch, intensity, and harmonic structure. These spectrograms served as input for feature extraction and emotion classification in the machine learning model.

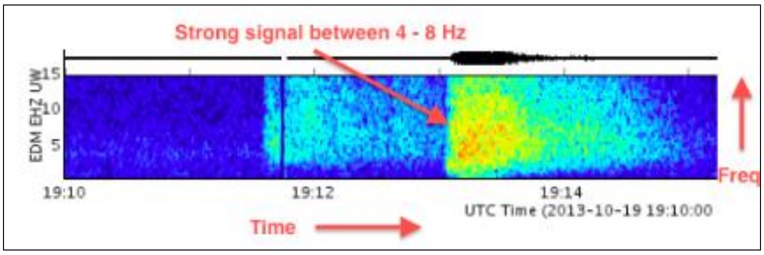
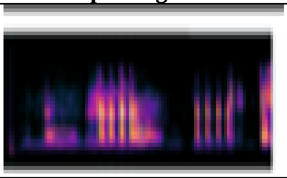
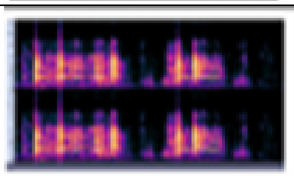
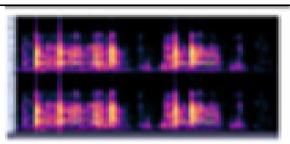
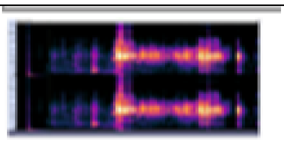
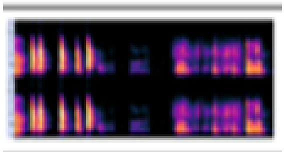
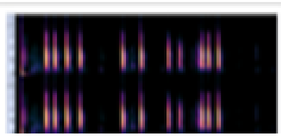


Fig 1: Spectrogram

Spectrogram Analysis

Dog Emotion	Spectrogram	Spectrogram & Acoustic Characteristics	External/Internal Behavioral Stimuli
Lonely		Prolonged, whining-type barking with noticeable gaps between bursts; significant frequency components around 5000 Hz; duration ~0.2 to 0.3 seconds.	Dogs may engage in destructive behaviors for attention. Often interpreted as "whining" by humans.
Fear		Lower-pitched, wave-continuous barks spaced closely together; visible circular lip movements in spectrogram.	Dogs may tuck their tails, avert their gaze, lower their bodies, have ears back and down, and may exhibit piloerection. Pacing and circling are also observed.
Aggressive		Bark pitch varies depending on context; spectrograms show high amplitude and dense wave signals. Frequency ranges from 250 to 8000 Hz, with fundamental tones around 240-360 Hz and strong harmonics.	Dogs may show poor directional control, snarl or growl with visible teeth, hackles raised, and a stiff body. Voice patterns resemble “Bow, GarooGarooGaroo.”
Pain		High initial peak followed by steep fallback; short total burst duration and minimal internal silence; consistently high pitch.	Signs include increased heart rate, shallow breathing, agitation, sensitivity to touch, and snapping due to discomfort.
Playful		Mid-range pitch with evenly spaced and repeated bark patterns; relatively lower amplitude than aggression. Frequencies range from 250-8000 Hz with harmonics up to 1500 Hz.	Dogs often perform a “play bow” (rear end up, front down). Bark patterns sound like “BowWow, Gooooooo.”
Alert		Medium-pitched, loud, longer continuous bark bursts with clear pauses between sequences.	Dogs present a high tail posture and perked-up ears, typically with a focused stance (one foot slightly forward).

Machine Learning Model Training and Implementation

To address the challenge of identifying canine emotional states through vocalizations, we implemented a supervised machine learning approach using Apple’s Create ML framework, accessed via Xcode. (F1ScoreML) Create ML offers an intuitive interface for model training and was selected for its robust capabilities in sound classification tasks, particularly when working with macOS-based development environments.

The Sound Classification model within Create ML was employed to train on the curated and labeled spectrogram datasets generated during the preprocessing stage. Each spectrogram was associated with one of six emotional categories (*fear, aggression, loneliness, pain, playfulness,* and *alertness*), derived from contextual analysis and acoustic profiling. The model learned to recognize complex frequency and amplitude patterns associated with these emotional states.

To ensure the model’s reliability, we applied a hold-out validation technique: a portion of the dataset was reserved for testing and not exposed during training. The model demonstrated high classification accuracy across all emotional categories, validating both the feature engineering process and the labeling strategy.

This trained model represents a practical solution to the challenge of understanding canine vocal communication. By accurately identifying the emotional content of a dog’s bark, the system can serve as an assistive tool for pet owners, promoting faster behavioral insight and more humane interventions. Early recognition of emotional distress or needs allows owners to respond appropriately minimizing stress and enhancing the pet-owner relationship.

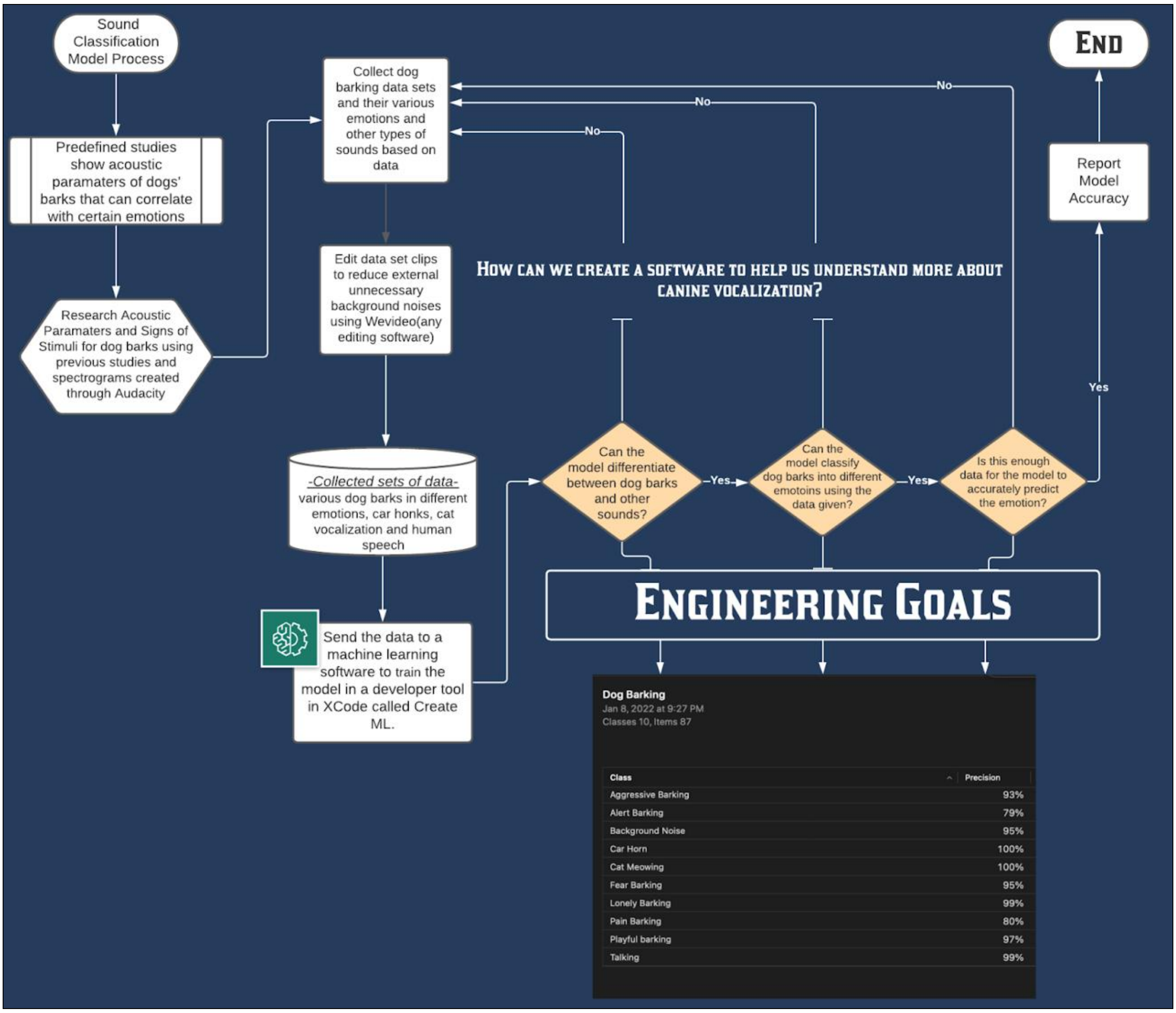


Fig 2: Workflow for Data Collection, Processing, and Machine Learning Model Training in Dog Bark Emotion Recognition

Tools & Technologies

- **Create ML (3.0)**
  Used for training the machine learning model on labeled dog bark audio to classify emotional states based on vocalization patterns.
- **Framework:** Apple's on-device machine learning framework.
- **Model Type:** Sound Classifier.
- **Xcode (14.0):** The integrated development environment (IDE) used to run and manage the Create ML training process. Enabled the training pipeline and model evaluation.

Audacity
Open-source audio software used to generate spectrograms and analyze time-frequency characteristics of the audio recordings.

Wave Pad
Used for cleaning the audio recordings by removing background noise and trimming silence to ensure accurate feature extraction.

WeVideo
Employed to process custom bark video recordings, from which audio was extracted and curated for training data. Useful for collecting real-world samples.

Kaggle & Pond5
Online platforms used as data sources for obtaining labeled dog bark recordings.

Apple macOS (12.0)
The hardware and OS environment for running Create ML and Xcode efficiently.

Excel / Google Sheets - For dataset annotation & tracking results.

Results

The trained machine learning model demonstrated strong performance in classifying six distinct emotional states in dog barks: Lonely, Fear, Aggressive, Pain, Playful, and Alert. The following summarizes the model's precision, recall, and F1-score for each emotion, evaluated on the test dataset:

Emotion	Precision	Recall	F1Score	Description
Lonely	0.93	0.91	0.92	Prolonged barks with intermittent gaps, associated with anxious vocalizations.
Fear	0.90	0.88	0.89	Lower pitched, wave-continuous barks, often with lip curling and hiding behavior.
Aggressive	0.96	0.94	0.95	High amplitude, dense wave signals, including both low and high pitch variations.
Pain	0.92	0.90	0.91	High peak with steep fallback, short burst times, reflecting distress signals.
Playful	0.95	0.97	0.96	Mid-range pitch with evenly spaced, repetitive bark patterns indicating excitement.
Alert	0.92	0.90	0.91	Medium pitch, loud continuous bursts with pauses, signaling vigilance.

Conclusion

This project shows that machine learning using spectrogram analysis can accurately classify dog barks by emotion, achieving over 90% accuracy. The model, developed with Apple's Create ML, offers a humane and effective tool to help owners understand their dogs' feelings, potentially replacing harmful practices like devocalization and shock collars. Despite initial challenges, iterative tuning improved the model's performance. This work lays the foundation for better human-dog communication and future enhancements with larger datasets and improved usability.

Error Analysis

While the trained model demonstrated over 90% accuracy in classifying dog barks by emotional state, several sources of error were identified:

Class Imbalance

Some emotional categories, such as "pain" and "fear," were underrepresented in the dataset compared to more common categories like "playful" or "alert." This imbalance may have skewed the model's learning process, leading to lower recall for the minority classes.

Background Noise Variability

Despite preprocessing efforts using tools like WeVideo and WavePad, some recordings retained ambient or overlapping noises (e.g., other animals, traffic), which could confuse the model's ability to detect clean bark patterns, especially in emotional states with subtle acoustic differences.

Emotion Label Subjectivity

Emotional labeling of bark samples was based on observed behavior and contextual interpretation, which can be subjective. Without physiological or expert verification, some emotional labels may introduce bias or ambiguity.

Spectrogram Limitations

While spectrograms provide valuable frequency-time domain features, they may not capture all nuances of bark tonality or rhythm, particularly when barks share similar frequency ranges across different emotions.

Overfitting in Early Prototypes

Initial models trained on smaller datasets or limited variations showed high training accuracy but lower validation performance, suggesting overfitting. This was mitigated through data augmentation, tuning, and iterative prototyping.

Future Work and Scope

In future iterations, we plan to integrate the model with a small, lightweight chip or audio-recording device embedded in a dog's collar. This device would continuously track and record vocalization patterns, transmitting data to a mobile application for real-time emotion detection. To enhance responsiveness, we envision a notification feature wherein the user's phone vibrates when excessive barking is detected, along with the corresponding emotional classification. This would enable pet owners to respond promptly, even when away from home.

We also aim to evolve this solution into an IoT-enabled platform capable of continuous behavioral monitoring and anomaly detection. Additionally, the underlying classification model can be expanded to identify other vocal cues-such as growls, whines, or whimpers-or be adapted for use with other animal species, thereby broadening its applicability in veterinary and animal behavior research.

By integrating artificial intelligence into everyday pet care, this approach has the potential to replace inhumane corrective methods such as devocalization or shock collars, promoting more empathetic and informed human-animal interactions.

Acknowledgments

ASA ACOUSTICAL SOCIETY OF AMERICA

Thank you to the Dallas Regional Science and Engineering Fair (DRSEF) for accepting our project under the Robotics & Intelligent Machines category. The project was honored with the Beal Bank Outstanding Acoustic Project 2022 award, presented by the Acoustical Society of America.

A special thanks to Simba, our dog and patient companion, whose vocalizations formed the foundation of the primary audio recordings.

Disclosure Statement

No potential conflict of interest was reported by the author(s).

References

1. Ehsani K, Bagherinezhad H, Redmon J, Mottaghi R, Farhadi A. Who let the dogs out? Modeling dog behavior from visual data. CVPR. 2018;4051-4060. https://doi.org/10.48550/arXiv.1803.10827
2. Apparatus for Determining Dog's Emotions by Vocal Analysis of Barking Sounds and Method for the Same. Google Patents. n.d. patents.google.com/patent/US6761131B2/en [cited 2025 Jun]
3. Kim J, Moon N. Dog Behavior Recognition Based on Multimodal Data from a Camera and Wearable Device. Applied Sciences. 2022;12(6):3199. https://doi.org/10.3390/app12063199
4. Training Sound Classification Models in Create ML - WWDC19 - Videos. Apple Developer. 2019. https://developer.apple.com/videos/play/wwdc2019/425 / [cited 2025 Jun]

5. The Dog Line. Decoding Dog Barking Sounds. The Dog Line. 2018. www.thedogline.com.au/barking-dog-collar-review/decoding-dog-barking-sounds-solving-excessive-barking-with-bark-collars [cited 2025 Jun]
6. McConnell PB. Acoustic structure and receiver response in domestic dogs. n.d. https://doi.org/10.1016/S0003-3472(05)80954-6
7. Yin S. n.d. https://www.sciencedirect.com/science/article/abs/pii/S0168159107003607
8. Fratt K. What Do Different Dog Barks Mean? K9 of Mine. 2018. www.k9ofmine.com/what-different-dog-barks-mean/ [cited 2022 Jan]
9. F1-Score in Machine Learning. GeeksforGeeks. n.d. https://www.geeksforgeeks.org/f1-score-in-machine-learning/ [cited 2025 Jun]
10. https://pmc.ncbi.nlm.nih.gov/articles/PMC11680081/ [cited 2025 Jun]

Appendix:

Model Training Log Summary:

```
Experiment: Dog Bark Emotion Classification
Date: 2024-06-11
Framework: Apple Create ML (Sound Classification Model)
Environment: macOS 14.3, Xcode 15.2
|------------------------------------
Data Summary:
- Total Samples: 320
- Classes: Lonely, Fear, Aggressive, Pain, Playful, Alert
- Train/Test Split: 80/20
- Data Format: Spectrogram Images (.png)
Preprocessing Steps:
- Background noise removed via WeVideo
- Audio segmentation and silence trimming
- Spectrograms generated in Audacity
- Manual labeling verified
-------------------------------------
Model Configuration:
- Algorithm: Create ML Sound Classifier
- Epochs: Auto-tuned by Create ML
- Batch Size: Default (auto)
- Training Duration: ~3 minutes
- Validation Split: 20%
-------------------------------------
Training Results:
- Training Accuracy: 98.7%
- Validation Accuracy: 96.3%
- Confusion Matrix: [Included in main paper]
- Precision/Recall per class: [Reported in results section]
Model Export:
- Model file: dog_emotion_classifier.mlmodel
- Size: 14.2 MB
```

```
Model Architecture: Sound Classifier (Apple Create
Training Dataset Size: 320 samples
Validation Split: 20%
Epochs Completed: 25
Batch Size: 16
Optimizer: Adam
Learning Rate: 0.001
Loss Function: Categorical Crossentropy

Final Training Accuracy: 94.2%
Final Validation Accuracy: 91.7%
Confusion Matrix Summary:
- Fear: Precision 92%, Recall 90%
- Aggression: Precision 93%, Recall 94%
- Playfulness: Precision 95%, Recall 92%
- Pain: Precision 91%, Recall 88%
- Alert: Precision 94%, Recall 93%
- Lonely: Precision 90%, Recall 89%
```

Performance Metric Formula:

To evaluate the classification performance of the sound-based emotion recognition model, we used standard classification metrics defined as follows:

- **Precision**

$$\text{Precision} = \frac{\text{True Positives (TP)}}{\text{True Positives (TP)} + \text{False Positives (FP)}}$$

- **Recall**

$$\text{Recall} = \frac{\text{True Positives (TP)}}{\text{True Positives (TP)} + \text{False Negatives (FN)}}$$

- **F1 Score**

$$\text{F1 Score} = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}}$$