Classifying Emotions in Puppy and Cat Vocalizations Using Spectral Features and Classical Machine Learning

Yidan Chen

Abstract

This project explores how machine learning can be used to classify the emotional states of dog and cat vocalizations based on spectral features. Dimensionality reduction and classical classification methods—including Principal Component Analysis (PCA), Linear Discriminant Analysis (LDA), and Logistic Regression—were applied. While classification performed well for more stable vocal traits such as breed, emotional state classification proved more challenging. The results suggest that richer data and more flexible models are needed to accurately capture emotional information in animal sounds.

Contents

1 Introduction

Understanding emotional communication in animals—especially domestic pets—is an emerging area in behavioral science and pet technology. Vocalizations from dogs and cats often reflect emotional states, and decoding these signals can enhance human-animal interaction and well-being monitoring.

This project began with a simple question: **Can we distinguish a puppy's emotional state based on its bark?** We started with 45 unlabeled bark spectrograms (each with 1025 features), but faced challenges due to high dimensionality and limited data. To address these issues, we expanded our analysis to a larger dataset of 400 cat vocalizations. We also explored classification on related variables such as breed.

This report presents our machine learning approach to the problem, summarizes key results and lessons learned, and suggests directions for future improvement.

2 Methods and Results

2.1 Data and Preprocessing

Puppy Dataset:
We used 45 audio samples of dog barking (.wav format). Since these recordings were originally unlabeled, we manually assigned each sample one of three emotional states—**aggressive, defensive**, or **normal**—based on empirical auditory interpretation. This manual labeling process was necessary for supervised learning but inherently subjective, introducing the potential for bias and labeling error.

Cat Dataset:
The cat meowing dataset consisted of 400 labeled audio samples. Each sample was annotated with two attributes: *Breed*: **Maine**

Coon or **European Shorthair** and *Emotional States*: **B: brushing, F: waiting for food** or **I: isolation in an unfamiliar environment**.

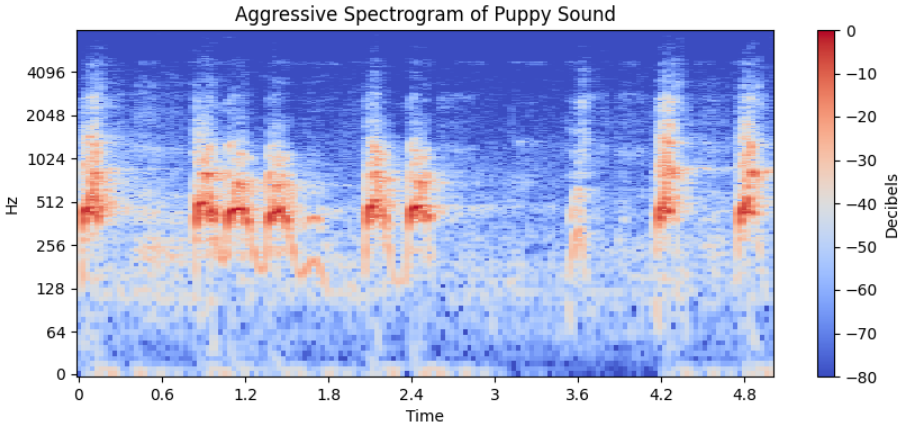

Figure 1: Spectrogram of a single sample

Preprocessing: We performed preprocessing using Python the `Librosa` library. Each waveform was transformed into a spectrogram which captures both time and frequency information. From these spectrograms, we extracted the mean spectrum by averaging the decibel-scaled frequency bins over time and saved them into a `CSV` file for following analysis.

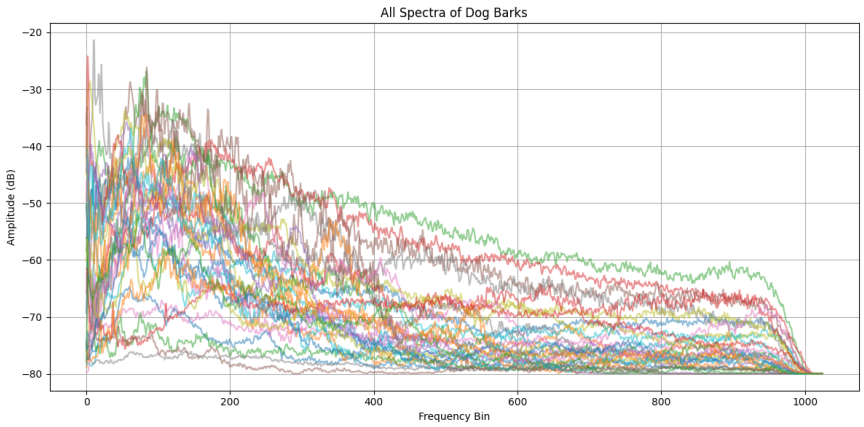

Figure 2: All Spectra of Dog Barks

2.2 Dimensionality Reduction and Classification

Initial PCA on Unlabeled Puppy Data

PCA was used to reduce 1025-dimensional feature space, aiming to capture the most variance. However, no clear cluster structure emerged from the principal component scatter plots, and classification models trained on PCA features performed poorly.

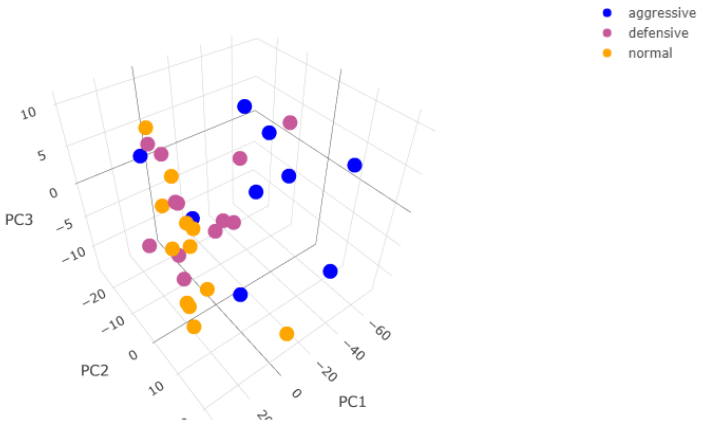

Figure 3: OCA-dog Emotion

Manual Labeling and LDA

The bark samples were manually labeled into three emotional categories. LDA was then applied. Although visualizations of the LDA-transformed space showed clear class separation, test set accuracy was low, with misclassifications frequent.

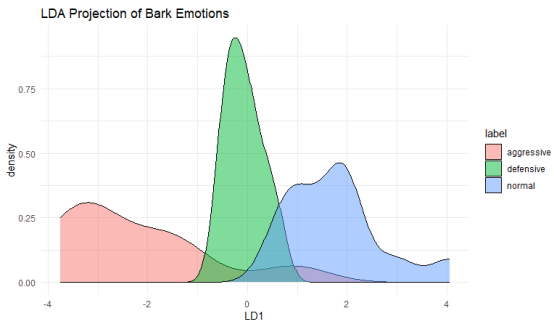

Figure 4: Raw LDA

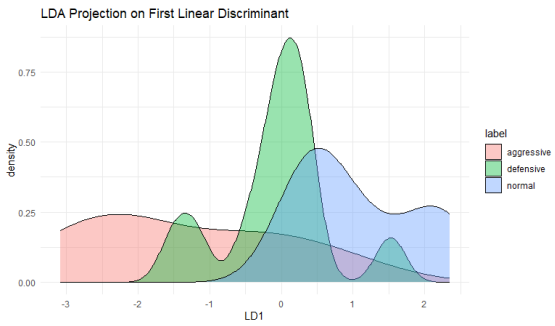

Figure 5: PCA+LDA

Analysis of assumptions showed skewed feature distributions and differing within-class variances, violating LDA's assumptions and explaining the poor performance.

PCA + LDA

To address LDA's sensitivity to violated assumptions, we used PCA as a preprocessing step. The PCA+LDA pipeline modestly improved classification accuracy but remained limited by the quality and size of the dataset.

Binary Logistic Regression on PCA-Reduced Features

Collapsing emotions into two groups (normal vs unnormal), we trained logistic regression on PCA features. This setup achieved the highest accuracy (around 75–80%), but at the cost of emotional granularity. Such a binary classifier lacks practical interpretability, as distinguishing "normal" from "not normal" offers little insight into behavior.

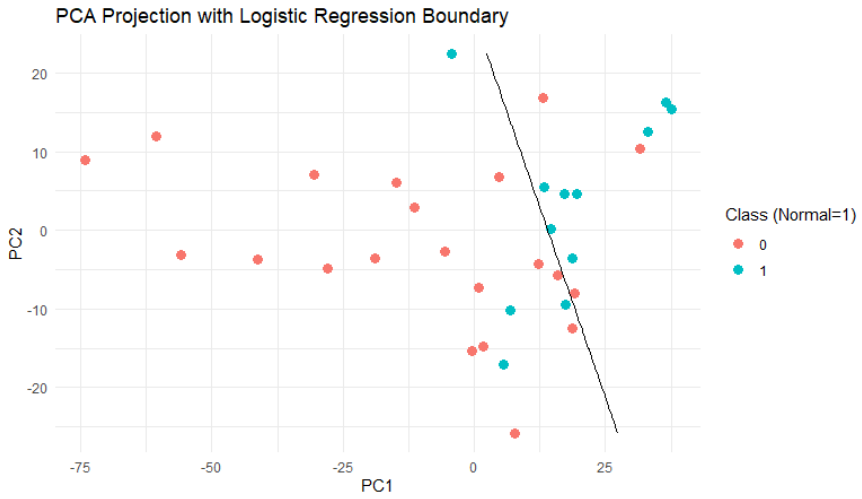

Figure 6: Binary Logistic Regression-Dog Barks

2.3 Extension to Cat Sound Dataset

With a larger dataset of 400 cat vocalizations, we replicated the pipeline. First, we tackled breed classification using PCA and LDA. Both models performed well, with accuracies exceeding 85%, show-

ing that vocal traits are effective for identifying breed differences.

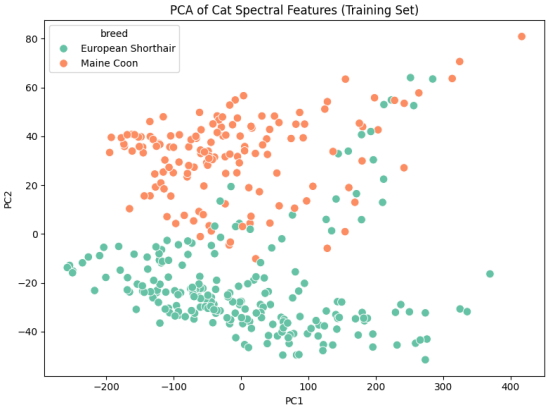

Figure 7: PCA-Cat Breed

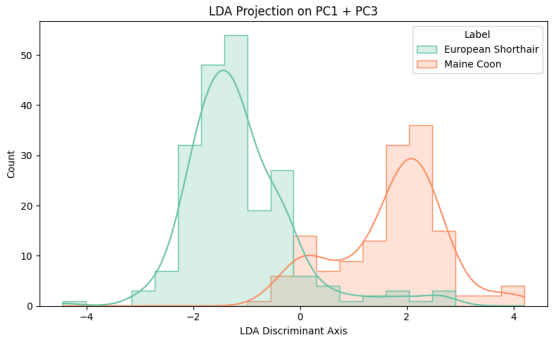

Figure 8: LDA-Cat Breed

We then applied the same approach to emotion classification. Despite the larger sample size, emotion classification accuracy dropped, indicating that emotion states is not as clearly encoded in vocal features as breed identity. The features that distinguished breeds didn't transfer well to distinguishing emotions.

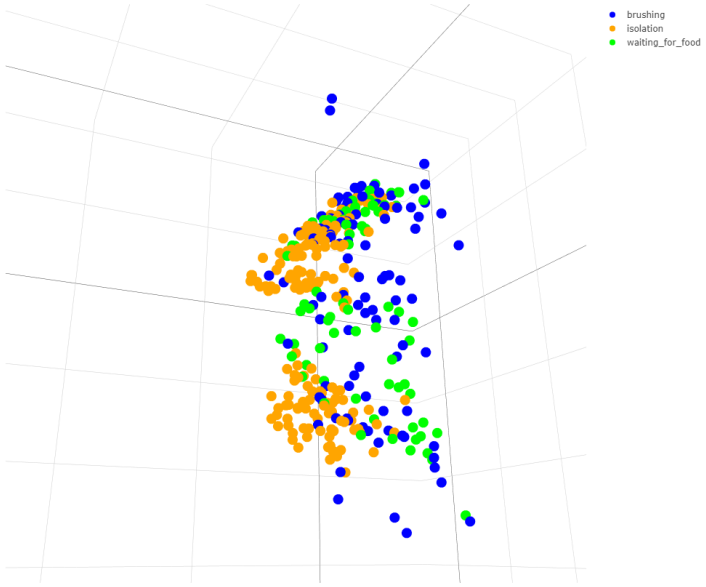

Figure 9: PCA Multinomial Logistic-Cat Meow Emotion

2.3.1 Nonlinear Modeling with Random Forest

To better capture the potential nonlinear relationship between spectral features and emotion, we trained a Random Forest classifier on the raw spectrogram data. we visualized its proximity matrix using multidimensional scaling (MDS). The resulting plot shows partial clustering within the `isolation`, while the `brushing` and `waiting for food` overlap in their vocal patterns.

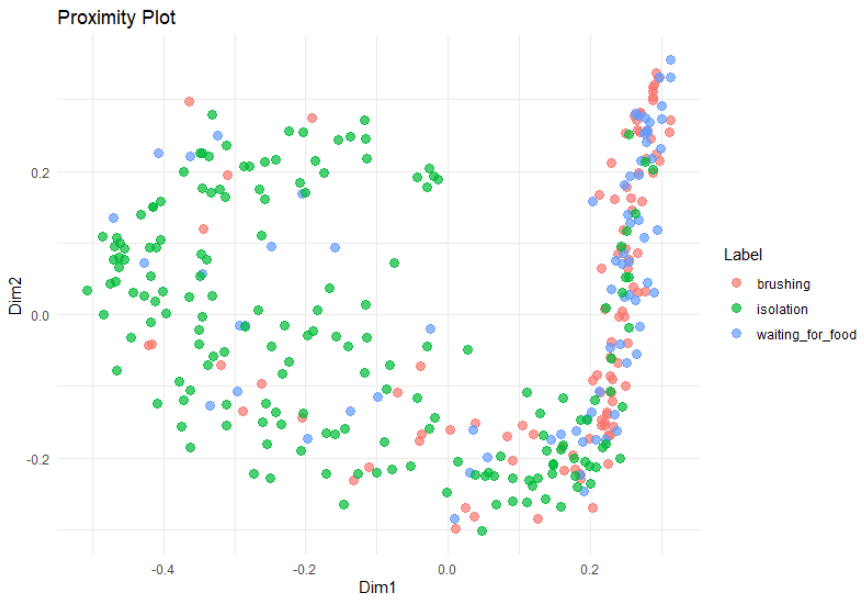

Figure 10: PCA Multinomial Logistic-Cat Meow Emotion

We further analyzed the top-ranked spectral features using partial dependence plots, which showed strong nonlinear responses across emotion classes. For example, the predicted probability for isolation dropped sharply above a specific threshold of the top-ranked feature. This kind of threshold-like behavior highlights Random Forest's ability to capture complex, nonlinear effects that linear models like LDA or logistic regression are unable to represent effectively.

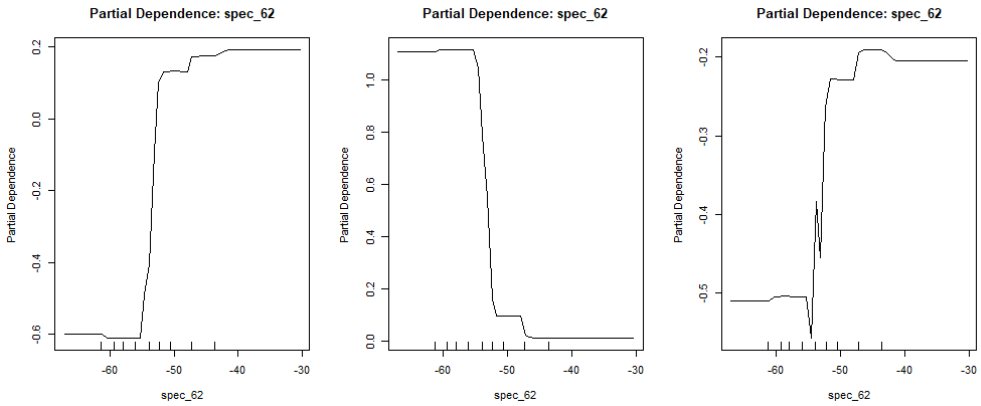

Figure 11: Partial dependence plots of `spec 62` for each class

2.4 Model Accuracy

Accuracy scores for each method and setting are summarized below:

Dog Emotion

Model	Accuracy
Multinomial Logistic (5 PCs)	61.5%
Multinomial Logistic (10 PCs)	61.5%
LDA (Raw Spectrum)	46.2%
LDA (PCA)	53.8%
Binary Logistic (Normal vs Other)	69.2%

Table 1

Cat Breed

Model	Accuracy
Multinomial Logistic (PCA)	94.3%
LDA (Raw Spectrum)	89.7%
LDA (PCA)	94.3%

Table 2

Cat Emotion

Model	Accuracy
Multinomial Logistic (PCA)	64.4%
LDA (Raw Spectrum)	52.9%
LDA (PCA)	65.5%
Random Forest	64.4%

Table 3

3 Analysis & Results

Our experiments highlight the difficulty of categorizing emotional content, as well as the difficulty of categorizing more stable vocal

features such as breed. While breed classification achieved high accuracy, emotion classification remains challenging, especially in small datasets.

For the **dog barking data**, the main limitations were the small sample size (N = 45) and the subjective nature of manual labeling, which also introduced class imbalance. Classical linear models like LDA performed poorly due to violated assumptions, such as equal within-class covariance and normality. Even after applying PCA to reduce dimensionality, classification accuracy remained low. A binary classification approach (normal vs. abnormal) using logistic regression yielded the best performance, but at the cost of reducing emotional subtlety. These results suggest that classical models may be insufficient for capturing subtle acoustic cues tied to emotional states, especially in high-dimensional, low-sample-size settings.

In contrast, the **cat dataset**—which included 400 samples with professionally annotated emotion and breed labels—enabled more robust analysis. Breed classification consistently achieved high accuracy in both linear and PCA-based models, suggesting distinct acoustic differences between Maine Coon and European Shorthair vocalizations. However, emotion classification remained difficult: accuracy plateaued around 65% across models, including PCA-based logistic regression and Random Forests. This may reflect the overlapping nature of emotional vocalizations such as brushing, food anticipation, and isolation.

Interestingly, **nonlinear modeling** with Random Forests provided only moderate improvements in cat emotion classification. Visualization of the proximity matrix revealed partial clustering (especially for the "isolation" class), and partial dependence plots indicated nonlinear threshold-like effects in key spectral features—patterns that linear models could not capture. This reinforces the idea that emotional expression in vocalizations may depend on complex, nonlinear interactions among spectral components.

4 Discussion and Conclusion

This project explored the feasibility of using spectral features and classical machine learning methods to classify emotional states in animal vocalizations—specifically dog barks and cat meows. While breed classification performed well, emotion classification proved to be significantly more challenging, reflecting the nuanced and less structured nature of emotional expression in sound signals.

Our findings highlight several key challenges:

- **Data limitations:** The puppy dataset was small and manually labeled, making it vulnerable to noise, class imbalance, and overfitting. Even the larger and more reliable cat dataset suggested that spectral features alone may not be sufficient to capture subtle emotional cues.
- **Model limitations:** Linear models such as LDA and logistic regression are straightforward to implement and interpret, but their performance declines when underlying assumptions (e.g., normality, equal variance) are violated. The limited improvement from Random Forests suggests that more sophisticated nonlinear models may be necessary to effectively capture emotional patterns in vocalizations.
- **Feature limitations:** Our approach focused on static spectral features, averaging across time and thus discarding potentially informative temporal dynamics. Features such as variation in pitch, rhythm, or modulation over time may play a key role in encoding emotional states.

4.1 Future work

- Expanding the dataset, with more high-quality, labeled emotional annotations.
- Incorporating temporal features to capture dynamic acoustic changes.
- Exploring more powerful nonlinear models, including convolu-

tional neural networks (CNNs), which can learn patterns directly from raw or minimally processed audio data.

- Incorporating different data modalities, such as video or images, could provide richer contextual information for emotion labeling and improve classification accuracy.

References

[1] A. Abzaliev, H. Pérez Espinosa, and R. Mihalcea. (2024, April 29). *Towards dog bark decoding: Leveraging human speech processing for automated bark classification*. Retrieved from https://arxiv.org/abs/2404.15123

[2] R. A. Lefèvre, C. C. R. Sypherd, and É. F. Briefer. (2025). Machine learning algorithms can predict emotional valence across ungulate vocalizations. *iScience*, 28(2), 111834. https://doi.org/10.1016/j.isci.2025.111834

[3] P. Pongrácz, C. Molnár, Á. Miklósi, and V. Csányi. (2005). Human listeners are able to classify dog (*Canis familiaris*) barks recorded in different situations. *Journal of Comparative Psychology*, 119(2), 136–144.

[4] L. Shen, M. J. Er, and Q. Yin. (2022). Classification for high-dimension low-sample size data. *Journal of Advanced Research*, 12(3), 45–58. https://doi.org/10.1016/j.jarr.2022.06.003