A Decoder of Domestic Cat Vocalizations

Haoheng Tang, Joanna Shen, Ziqian Liao

1. Introduction

This study investigates the feasibility of utilizing domestic cat vocalizations to classify their emotional states. A significant challenge in pet ownership lies in accurately interpreting the needs and affective conditions of animals, a task that proves particularly complex in feline species due to the high variability in their acoustic expressions. Individual cats may employ distinct vocal signals to convey analogous meanings, introducing ambiguity in human-animal communication. Given the nuanced and often context-dependent nature of feline vocalizations, this research explores the application of artificial intelligence to enhance human understanding of cat emotions, thereby facilitating improved care and welfare.

2. Methodology

2.1 Hypothesis

We hypothesize that feline emotional states—such as pleasure, anger, or pain—can be reliably inferred solely from vocalizations, independent of supplementary behavioral or visual cues. If validated, this would demonstrate that machine learning models can classify affective states in domestic cats using acoustic features alone.

2.2 Research Objective

The primary objective of this study is to develop an artificial intelligence model capable of accurately classifying feline emotions from vocal signals. Given an input audio sample, the model will predict the most probable emotional state, outputting a discrete label (e.g., "Anger," "Happiness," "Resting"). This approach aims to provide a scalable and non-invasive tool to improve human understanding of domestic cat welfare needs.

2.3 Dataset

For model training and evaluation, we utilize the *CatSound* dataset, originally developed by Pandeya and Lee (2018). This dataset comprises 5,922 audio samples of domestic cat vocalizations, sourced from publicly available videos on YouTube (https://www.youtube.com/) and Flickr (https://www.flickr.com/). The samples are annotated into 10 distinct emotional and behavioral categories:

1. *Anger*
2. *Defense Behavior*
3. *Fighting*
4. *Happiness*
5. *Hunting Desire*
6. *Mating*

7. *Calling for Its Mother*
8. *Pain*
9. *Resting*
10. *Warning*

Each category represents a feline emotional state, and is used as a class label for our model (Figure 1).

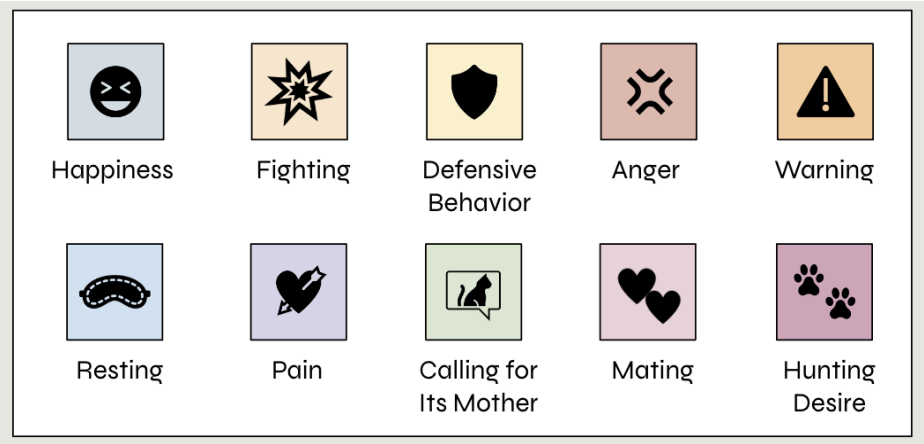

Figure 1. Categories of Cat Vocalizations

During preprocessing, we converted the audio from .mp3 format to .wav to comply with Python packages. We also converted the stereo input (2 channels) into mono (1 channel). There is no missingness in our dataset.

2.4 Exploratory Data Analysis

Our dataset consists of 5,922 audio files. The durations of the audio files vary significantly, ranging from as short as 0.313 seconds to as long as 16.797 seconds. This variability reflects the natural differences in cat vocalizations and adds complexity to the task of emotion classification. Moreover, the distributions of audio lengths differ across the 10 emotional classes. Histograms (Figure 2) illustrating these distributions show that certain emotions are more likely to be expressed through shorter or longer vocalizations, indicating potential patterns worth investigating further.

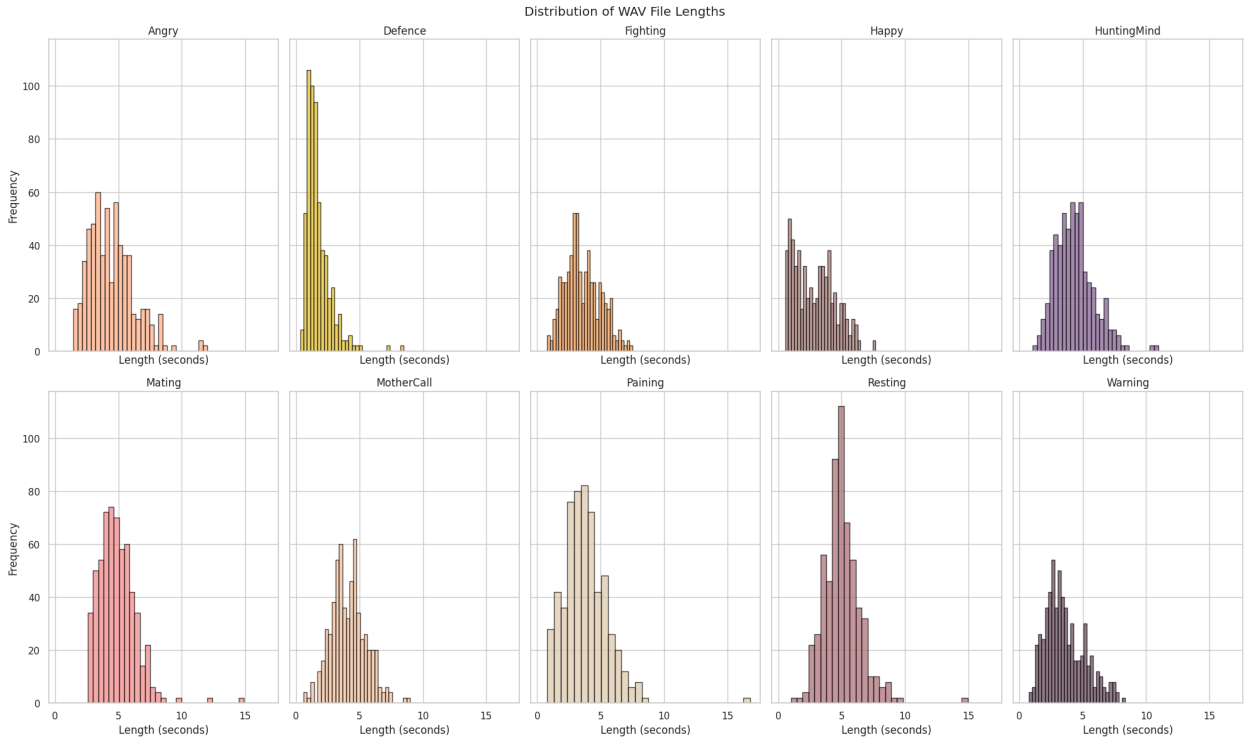

Figure 2. Distribution of Audio Length

An important characteristic of the dataset is its class balance. The 10 emotional categories are evenly represented, ensuring that the classification model is not biased toward any particular emotion. This balance is visually confirmed through the following bar graph (Figure 3), which shows roughly equal counts for each class.

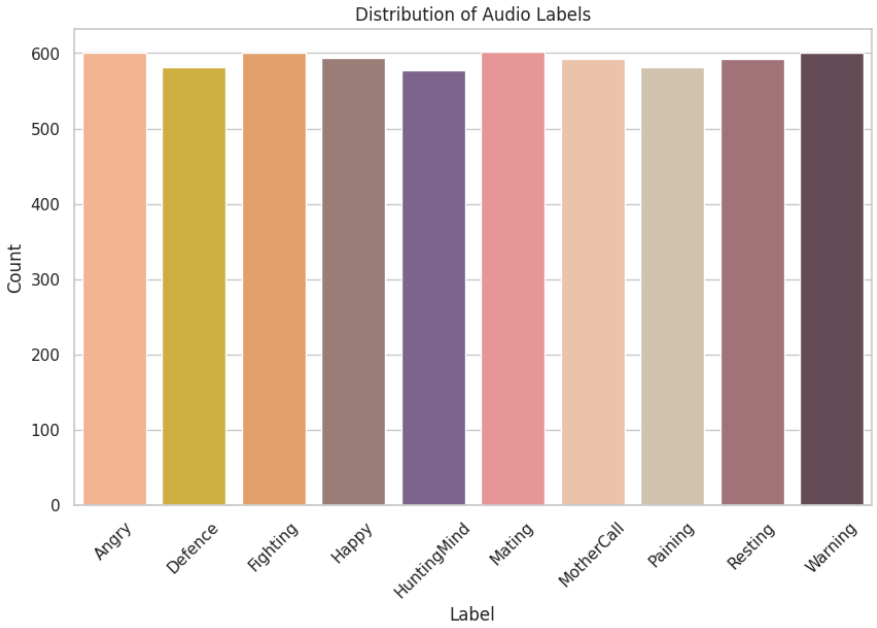

Figure 3. Distribution of Audio Labels

Our analysis identifies a critical challenge: substantial intra-class variability within individual emotion categories, juxtaposed with limited inter-class distinction in certain cases. This phenomenon is particularly observable in spectrographic representations (Figure 4). For instance, while the left two spectrograms exhibit marked visual dissimilarities, both correspond to vocalizations classified as Happiness. Conversely, the rightmost spectrogram, which bears a striking resemblance to the Happiness samples in the middle, is in fact associated with Pain. These observations indicate that while feline vocalizations demonstrate considerable diversity in the acoustic manifestation of a single emotional state, overlapping features across distinct emotions may impede reliable classification. This underscores the inherent complexity and nuance in decoding affective states from domestic cat vocal signals.

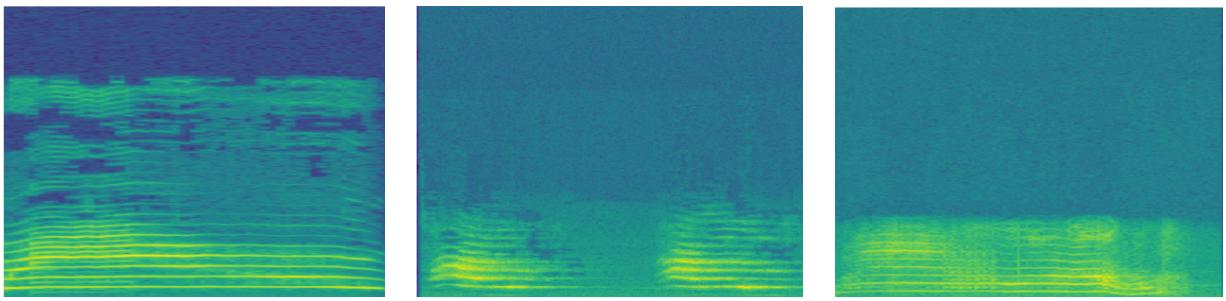

Figure 4. Spectrograms of Audios from Different Categories

2.5 Modeling Approach

Given the task of classifying the emotional state based on cat sound, we decided on combining an audio feature extractor and a Convolutional Neural Network (CNN) classifier to build our model.

Feature engineering has proved to be a crucial part of our task. We implemented different feature engineering approaches, ranging from a more traditional one to a transformer-based audio encoder, and compared their impact on model performance.

Our best feature engineering approach includes the computing of:

1. *Zero Crossing Rate*: Measures the rate at which audio signal changes from positive to negative or vice versa, implying the smoothness of the sound.
2. *Chroma Frequencies*: Measures the harmonic content in the sound.
3. *Mel-Frequency Cepstral Coefficients (MFCCs)*: Captures the timbre and texture of the audio.
4. *Spectral Rolloff*: Measures the distribution of total energy, implying the brightness of a sound.
5. *Spectral Bandwidth*: Quantifies the width of the band of frequencies that contain most of the energy of the signal.
6. *Root Mean Square Energy (RMS)*: Measures the signal's power, reflecting its loudness.

7. *Mel-Spectrogram*: Numerically represent the sound with frequencies converted to the Mel scale, a scale that approximate human ears' response to frequencies.

Such computing is implemented through a pipeline with the use of librosa, a Python module for audio and music processing. The resulting vectors were transposed and averaged over time frames to reduce the dimensionality while retaining the most significant characteristics that represent the audio sample. Finally, we concatenated these compressed statistics to form an array representative of the audio file. The array was then taken as the input to our CNN classifier to make predictions. The diagram below illustrates the pipeline of our model (Figure 5).

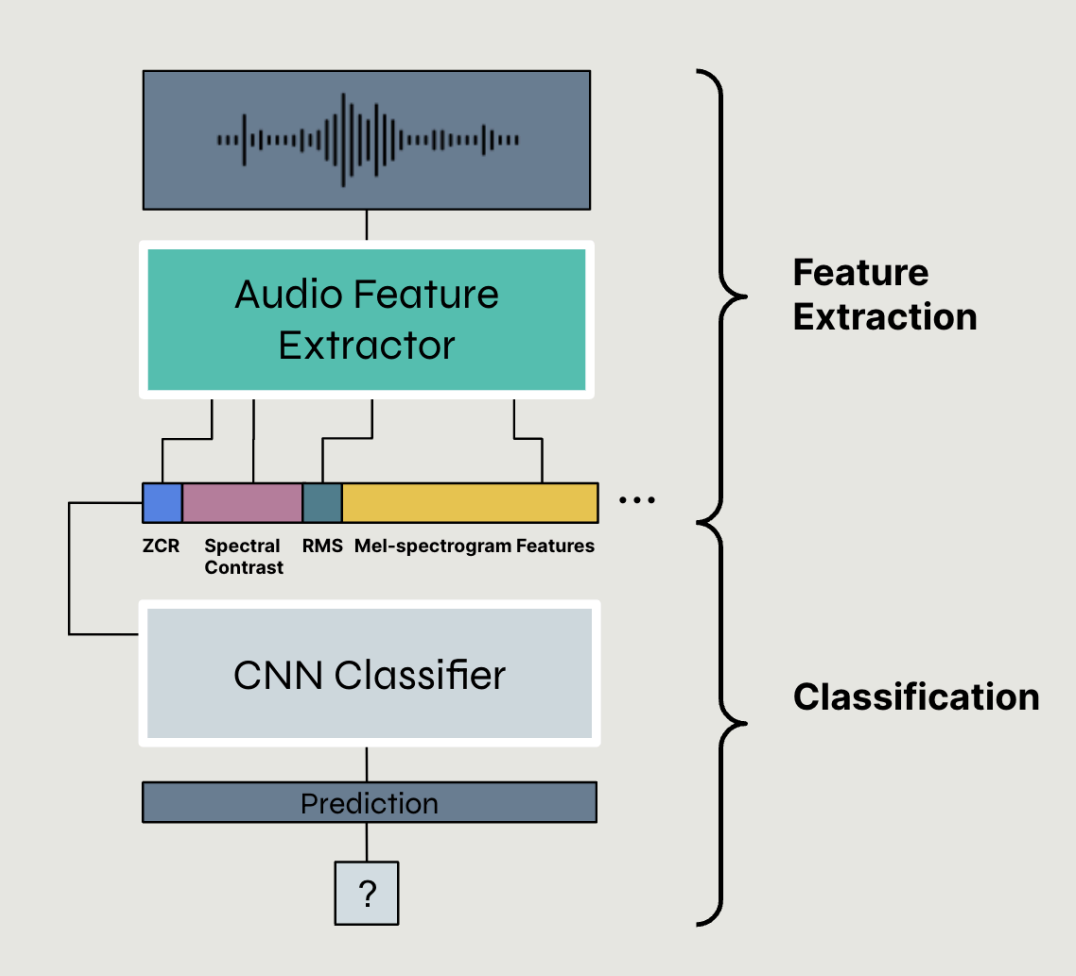

Figure 5. Model Architecture

3. Evaluations

The model was evaluated on a test set comprising 1,185 samples, achieving an overall classification accuracy of 85.15%. As detailed in Table 1, performance varied significantly across emotion categories. The model demonstrated particularly strong performance in classifying Fighting and Resting vocalizations, with accuracy rates approaching 93% for both categories. Conversely, classification of Happiness vocalizations proved most challenging,

yielding an accuracy of approximately 77%. This reduced performance aligns with the observed high intra-class variability within the Happiness category, as illustrated in Figure 4.

Class	Anger	Defensive Behavior	Fighting	Happiness	Hunting Desire	Mating	Calling for its mother	Pain	Resting	Warning
Accuracy	0.79	0.88	0.93	0.77	0.86	0.81	0.90	0.83	0.93	0.83

Table 1. Accuracy Rates by Class

The confusion matrix (Figure 6) is a structured table that summarizes the outcomes of our classification task, offering a comprehensive view of the model's predictive capabilities. It highlights both its effective and deficient areas of predictions and helps us understand the bias of our model.

The confusion matrix reveals strong model performance, as evidenced by the prominent diagonal pattern indicating correct classifications across most categories. However, systematic misclassifications occur between the Happiness and Pain sounds, as well as between the Warning and Mating sounds. These classification errors likely stem from acoustic similarities between the respective vocalization types: specifically, between the "meow-meow" pattern characteristic of Happiness and the "miyoou" pattern associated with Pain, and between the "ko-ko-ko" Warning vocalization and the "gay-gay-gay" Mating call (Pandeya & Lee, 2018). These findings are consistent with our earlier observations from exploratory data analysis (Figure 4), which similarly highlighted the acoustic overlap between these categories.

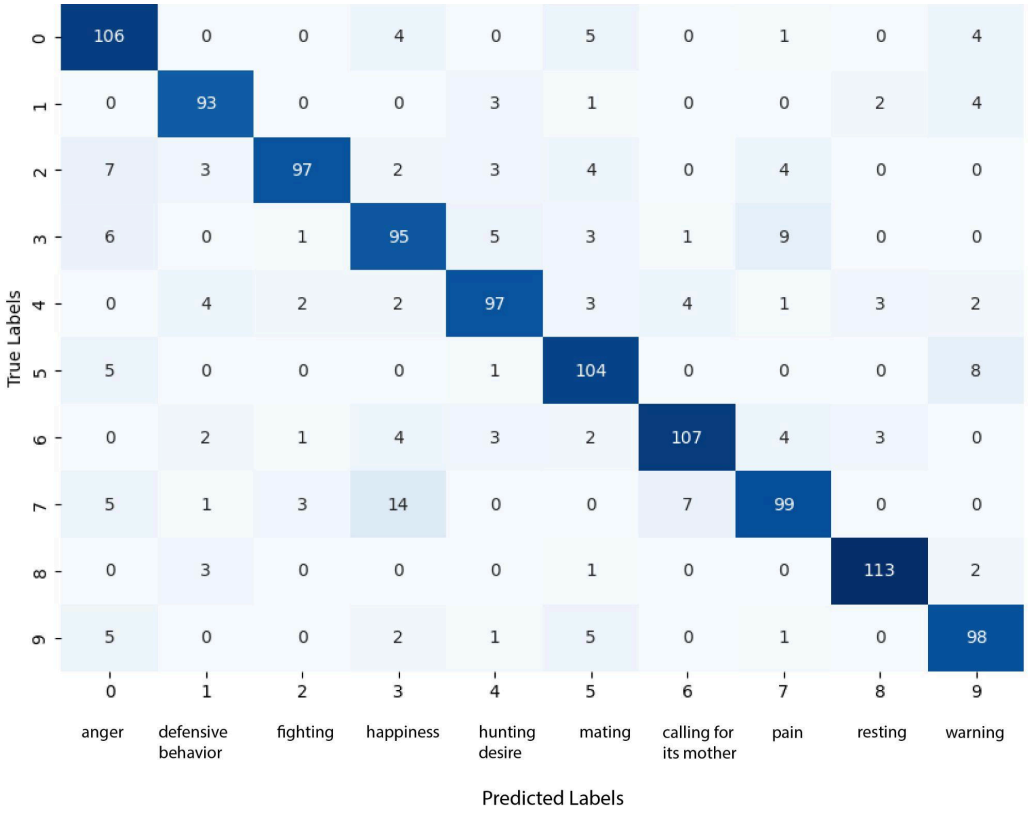

Figure 6. Confusion Matrix

4. Conclusions

This study demonstrates that an AI-based decoder of domestic cat vocalizations can serve as an effective, non-invasive tool for assessing feline affective states and welfare needs. Our results indicate that artificial intelligence models can classify emotional states from vocal signals with satisfactory accuracy, suggesting promising applications in veterinary behavioral science and pet care.

Several directions emerge for future research:

- **Dataset expansion**: Enhanced model generalizability necessitates a more extensive and diverse dataset. Future work will focus on systematic collection and annotation of additional samples from both digital repositories and real-world environments.
- **Model refinement**: The current architecture requires improvement in distinguishing between acoustically similar but emotionally distinct vocalizations. Alternative architectures, such as transformer-based models, may offer superior performance compared to the current CNN-based approach.
- **Practical implementation**: To facilitate the transition from experimental research to practical implementation, further technological development is required. Current efforts focus on creating a mobile and web-based application platform capable of processing feline vocal inputs and generating corresponding emotional state classifications. This proposed application would enable pet owners to better interpret their cats' affective states for improved welfare monitoring, and assist individuals in understanding unfamiliar cats' emotional expressions.

These developments would establish a more robust foundation for computational analysis of feline vocal communication while advancing tools for animal welfare assessment.

Reference

[1] Pandeya, Y. R., Kim, D., & Lee, J. (2018). Domestic cat sound classification using learned features from deep neural nets. *Applied Sciences*, *8*(10), 1949.

[2] Pandeya, Y. R., & Lee, J. (2018). Domestic cat sound classification using transfer learning. *International Journal of Fuzzy Logic and Intelligent Systems*, *18*(2), 154-160.