DCATVD: A LARGE-SCALE LONGITUDINAL DATASET FOR ANALYSIS OF DOMESTIC CAT VOCAL DEVELOPMENT

Anonymous authors
Paper under double-blind review

ABSTRACT

In longitudinal studies of animal vocalizations, machine learning models enable a deeper analysis and understanding of developmental patterns in animal communication and vocalizations. However, longitudinal studies remain constrained by the limited availability of high-quality, large-scale datasets. To address this gap in feline bioacoustics, this paper introduces the Domestic Cat Age Transition Vocalization Dataset (DCATVD), a large, robust dataset containing vocalizations from 135 individual domestic cats across 12 different breeds with precise metadata (exact birth-date, breed, age group, and individual cat ID). Our thorough analysis of this dataset reveals novel discoveries on how essential vocal variables develop across the lifespan of cats, such as shifts in dominant vocalization type from *purr* to *meow* and the increase of specific precise vocalization units (Elemental Cat Meow Units). This paper opens an avenue to advance bioacoustic representation learning through an accessible dataset and a reproducible pipeline.

1 INTRODUCTION

As cats age, they face changes in vocal anatomy (Tavernier et al., 2020), cognition (Denenberg et al., 2024), and behavior (Bellows et al., 2016). Analyzing these age-related changes in cat vocal communication is crucial to understand these aspects of feline health. Although there are more than 600 million domestic cats in the world (Kays et al., 2020), current research on feline bioacoustics is limited by the lack of large-scale longitudinal datasets that track individual cats throughout multiple stages of life while maintaining consistent age and breed metadata (Tang et al., 2024). The majority of existing studies on domestic cat bioacoustics are conducted in controlled laboratory environments with small sample sizes, making it difficult to study long-term vocal development in naturalistic settings. This highlights the need for more representative, larger datasets that will allow for these longitudinal studies on cat vocalizations. To address this need, this study investigates how the vocal repertoires of cats differ across their lifespan and between breeds by using advanced computational methods inspired by a similar study conducted on dog vocalizations by Lekhak et al. (2025).

We hypothesize that cats will exhibit changes in their vocal repertoires throughout their lifespan, just as humans show early vocal development and changes in language representation in the brain (Kent & Murray, 1982; Szaflarski et al., 2006). Reasonably, other animal species have also shown developmental changes in their vocal repertoires: mice use acoustic cues such as syllables and sequences to determine age of the caller (Grimsley et al., 2011), and marmoset monkeys, similar to humans, have prolonged vocal development that remains flexible until adulthood (Gultekin et al., 2021). Longitudinal vocal development research across different species remains a very active area of study and has the potential to uncover essential details of human language evolution (Fishbein et al., 2019). To explore this, we created one of the largest longitudinal datasets to date on feline age-related and breed-related vocalization changes. This dataset was formed by collecting videos of cats from 12 different breeds (detailed in Section 2.1.1), across six different life stages (defined by The American Animal Hospital Association (AAHA) and American Association of Feline Practitioners (AAFP) (Quimby et al., 2021)) using exact birth-dates from social media platforms like YouTube and Instagram, and applying audio processing and machine learning techniques to extract, process, and analyze cat calls. Additionally, we compare our resulting dataset with the Feline Age Predic-

Reviewers: please read the Reviewer Guidelines (iclr.cc/Conferences/2027/ReviewerGuidelines) and the AI Policy for Reviewers (iclr.cc/Conferences/2027/AIPolicyForReviewers). If you used AI to expand, edit, or polish your review, please provide the input text to the LLM. Better still, consider skipping the LLM and submitting your original text: we, and the authors, are much more interested in your unedited thoughts than in what an LLM has to say. AI-assisted or not, you are putting your name and reputation behind your review: LLM-generated falsehoods, hallucinations or misrepresentations are subject to disciplinary action, which may include desk-rejecting all papers you have authored.

tion Dataset (van Toor et al., 2025) and find that we improve on their dataset with 11.5 times more vocalizations and a more scalable, robust pipeline for longitudinal cat vocal development studies.

Building upon the computational methods established by Lekhak et al. (2025), this dataset supports research on feline vocal development and age-related changes. Understanding how the vocalizations of domestic cats change over their lifespan and across breeds reveals essential elements of their aging and social communication. This kind of research is essential given that cats are an integral part of human society (Lyons, 2014; Lipinski et al., 2008), having lived alongside humans for thousands of years. Ultimately, this research provides crucial insight into cat communication and vocal development, revealing how computational methods can be used to uncover these complexities in cats and possibly other animals.

We adopt terms similar to those used by Lekhak et al. (2025), which we define below for clarity throughout this paper:

- **Meow Unit (MU)**: An individual continuous vocal sound produced by a cat, representing a distinct vocal utterance. MUs are delineated from surrounding non-vocal segments or adjacent MUs based on acoustic properties such as significant energy dips and low variability (detailed in Section 2.1.4).
- **Elemental Cat Meow Unit (ECMU)**: An acoustically distinct sound unit corresponding to a 25ms frame within an individual Meow Unit (MU). These units are classified into a set of fundamental ECMU types derived by clustering Mel-frequency cepstral coefficient (MFCC) features from such frames (detailed in Section 2.3). Each Meow Unit is represented as a series of ECMU type labels assigned to its constituent frames.

Our key contributions are:

- We introduce a thoroughly curated Domestic Cat Age Transition Vocalization Dataset (DCATVD), which includes accurate metadata for 135 individual cats from 12 different breeds. This metadata includes precise birth-date, breed, age group, and individual cat ID in accordance with each of the 9,135 meow units (totaling 4.42 hours). This is the largest open-source cat dataset of this type, designed to enable novel and robust research regarding feline vocal development and bioacoustic representation learning.
- We put forward a reproducible pipeline for dataset creation involving precise metadata allowing us to curate the largest breed and age-accurate feline dataset of its kind [1].
- We present a comprehensive longitudinal study utilizing this rich dataset as our primary analytic contribution. This study reveals novel findings on changes in feline vocal patterns across ages and development. These findings include changes in Meow Unit call-type and Elemental Cat Meow Unit (ECMU) distribution across six life stages and twelve breeds (see Section 3).

2 METHODS

2.1 DATASET CREATION

Social media platforms offer vast amounts of audiovisual data that can be used to construct large-scale datasets (Rahman & Reza, 2022). Cat owners on platforms such as YouTube and Instagram post consistent videos of their pet cats over prolonged periods of time spanning from months to years, with many including the age or birth-date of their cat in video or channel descriptions, making these platforms ideal sources for data collection. We improve upon the dataset collection steps detailed in Lekhak et al. (2025) and construct our dataset using the data from these platforms with the following reproducible pipeline: (i) Channel Discovery and Selection, (ii) Video Filtration and Age Determination, (iii) Audio Preprocessing, and (iv) Meow Unit Segmentation.

[1] Our full dataset and code will be available upon acceptance. Anonymous data samples and code available for review here: `https://anonymous.4open.science/r/DCATVD-6F3F/`

2.1.1 Channel Discovery and Selection

Our dataset consists of videos gathered from the social media platforms YouTube and Instagram. We began with a search for relevant videos on YouTube using a list of targeted queries constructed from a combination of cat ages and breeds (e.g., "Siamese cat turned X years old"), selecting the top 20 most popular breeds of 2025 by number of registrations according to Governing Council of the Cat Fancy (2026). Additionally, we translated these queries into the native languages of the 12 countries with the highest YouTube user bases according to World Population Review (2026) to encompass a broad range of cat-owning communities. Queries generated in this manner allowed us to gather a large number of channels from across YouTube that are likely to publish videos of a specific cat and mention the cat's age and breed.

Next, we applied strict automated filtering to the first-pass videos in order to remove irrelevant channel candidates, excluding channels that did not reference age-related indicators or cat breed, channels with less than 10 published videos, and channels containing specific irrelevant keywords (e.g., "compilation," "facts," "for sale," or "guide"). To further eliminate poor-quality candidate channels, we used the Llama 3.1 8B model (Grattafiori et al., 2024) to analyze video titles and ensure that age references were associated with a single identifiable pet cat (e.g. "Does the video title focus on one particular cat? Does the video title indicate the age range of the cat?"). We validated our channel-filtration methods on a sample of 200 randomly selected channels, and achieved a 95.5% accuracy in eliminating channels containing no breed or age-related information.

Finally, we manually verified the validity of the remaining channels. Valid channels included those that specify (i) a focus on only one cat, (ii) the breed of the cat, (iii) the birth-date of the cat, and (iv) an upload range spanning over one year. Following this round of data collection, we focused our evaluation on the six breeds with the largest amount of data available on YouTube (Bengal, British Shorthair, Maine Coon, Persian, Ragdoll, Siamese). Alongside these breeds, we chose to include six additional breeds (Abyssinian, Birman, Devon Rex, Siberian, Sphynx, Tonkinese) in order to offer a more robust perspective on cat breeds in our analysis despite their more limited data.

To supplement our dataset and average out our video distribution over the main six breeds, we gathered 25 additional channels from Instagram using similar breed and age-related queries. This filtering resulted in 156 individual cat channels, with 131 cats from YouTube and 25 cats from Instagram.

2.1.2 Video Filtration and Age Determination

Following the selection of high-quality channels from YouTube and Instagram, we processed the videos from each channel to curate a set of all videos containing relevant feline vocalizations. To exclude videos irrelevant to the candidate cat, we employed CLIP-ViT, a neural network trained on (image, text) pairs, to detect the presence of a cat in the thumbnail image of each video (Radford et al., 2021). We eliminated any video with a confidence score of under 0.50 using the prompt "image of a cat."

We then employed PANNs (Kong et al., 2020) for sound event detection (SED) to filter out videos with an absence of feline vocalization. Trained on the AudioSet dataset (Gemmeke et al., 2017), the PANNs CNN14 model allowed us to verify the presence of cat vocalization events within each video. Using a threshold of 0.40 with the labels "Cat", "Meow", "Purr", and "Hiss", we achieved a 98% accuracy in detecting videos containing at least one cat vocalization on a sample of 100 videos. This left us with a total of **4,688 videos** containing one or more cat vocalizations from across **135 individual cats**. 21 of the original 156 cat channels were eliminated due to an absence of feline vocalizations within their videos.

The age of the cat within each video was calculated based on a set birthday or age identified in at least one video or channel description and cross-referenced with the upload date of each video. Following Vogt et al. (2010), cats were categorized into six age groups: Kitten (K; birth-6 months), Junior (J; 7 months-2 years), Prime (P; 3-6 years), Mature (M; 7-10 years), Senior (S; 11-14 years), and Geriatric (G; 15 years and older). Each video contains meticulous metadata including the individual cat ID, birth-date, breed, age, and age group of the vocalizing cat.

2.1.3 AUDIO PREPROCESSING

In our initial audio preprocessing phase, we isolated feline vocalizations from raw audio clips to prepare for downstream SED. We used the language queried audio source separation model, AudioSep (Liu et al., 2022), configured with text prompt "cat." It performed well for our applications, minimizing both the removal of feline vocalizations and the amplification of background noise.

We performed an evaluation of AudioSep on a random sample of 195 videos from our finalized dataset. Two testers were given corresponding original and denoised versions of these videos and assigned them a denoising rating from 1-5. Ratings were defined as follows: 1: complete cut of all cat vocalizations, 2: partial cut of cat vocalization, 3: no noticeable change, 4: some noise remained but no cat vocalization cut, and 5: removed all noise and left all cat. Results can be seen in Appendix A. We considered error to be the percentage of removed or partially cut vocalizations, ratings of 1 and 2, as vocalization cutting was our most relevant issue, given our dataset's size. Under this metric, AudioSep's averaged percent error was 8.97% with an inter-evaluator agreement of 90.76%.

2.1.4 MEOW UNIT SEGMENTATION

Our analysis of feline vocalization development relies on two representations of cat-call events: (i) Meow Units (MUs), and (ii) Elemental Cat Meow Units (ECMUs). Following the denoising step, we isolated bounded cat vocalization events using the feline-specific SED model introduced by Anonymous (2027). The SED model uses a frozen BEATs encoder with a recurrent temporal decoder and was trained using a semi-supervised method that combined strongly labeled cat vocalizations with synthetic, weakly labeled, and unlabeled audio. The strong-label dataset contained 1,353 clips with 3,165 verified cat-call boundaries, and the model achieved an event-level F1 score of 87.4% (Anonymous, 2027). We then applied the trained SED model to the denoised audio to identify temporal regions containing potential cat calls. Because a detected vocalization region can contain multiple individual vocal events, we applied the amplitude-based segmentation algorithm introduced by Lekhak et al. (2025) to obtain individual MUs.

We applied several filtering stages to retain only clean and reliable feline vocalizations. First, we removed MUs that did not receive a PANNs confidence score of at least 0.10 for any cat-related sound event class (*cat*, *meow*, *purr*, or *hiss*). We then excluded MUs longer than 6 seconds, as feline vocalizations most commonly fall within a substantially shorter duration range (Schötz et al., 2024). Finally, we trained a binary feline-vocalization classifier to distinguish MUs containing a feline vocalization from segments consisting primarily of background noise, human speech, or other irrelevant sounds.

The binary classifier was trained on a set of 4,815 MUs. Of these, 4,351 were verified as containing feline vocalizations, including the 4,250 MUs manually curated for the subsequent call-type classification task and an additional 101 vocal MUs, while 464 MUs were labeled as background or otherwise non-vocal material. The 4,250 call-type training examples were initially obtained by applying the call-type classifier adopted from Anonymous (2027) to our extracted MUs, followed by manually reviewing and relabeling the resulting samples. Two expert annotators independently listened to the MU samples and established the final labels by agreement. The resulting binary classifier was evaluated on a held-out test set of 963 MUs. As shown in Appendix B, the binary feline-vocalization classifier achieved **98.1%** overall accuracy, with an F1 score of 99.0% for feline vocalizations and 90.9% for background noise.

Only MUs classified as containing feline vocalizations were retained for analysis. The complete filtering pipeline yielded **9,135 individual MUs** from **135 cats**, corresponding to **4.42 hours** of vocalization audio. The retained MUs have a mean duration of **1.30 seconds**.

Table 1 details the breakdown of our final dataset. Appendix C contains the exact counts of individual cats and their corresponding MUs for each of the twelve breeds and six age groups. Our dataset beats all previous age-related feline vocalization datasets on size and metadata-precision. Compared to previous state-of-the-art (SOTA) cat vocalization datasets, the DCATVD boasts significantly more data sourced from a large number of individuals and is easily reproducible (Ludovico et al., 2021; van Toor et al., 2025; Schötz et al., 2024).

Table 1: Distribution of individual channels, videos, MUs, and total MU duration across all breeds.

Breed	Individuals	Videos	MU count	MU duration (hr)
Bengal	21	1,205	2,254	1.13
British Shorthair	15	532	1,000	0.48
Maine Coon	16	260	716	0.32
Persian	16	646	626	0.25
Ragdoll	27	777	1,922	0.96
Siamese	18	581	1,366	0.63
Abyssinian	3	119	210	0.12
Birman	3	189	311	0.16
Devon Rex	4	60	138	0.06
Siberian	6	203	465	0.24
Sphynx	2	93	106	0.05
Tonkinese	4	23	21	0.02
Total	**135**	**4,688**	**9,135**	**4.42**

2.2 Classification of Meow Units into Cat Call-Types

In order to categorize each MU into a cat call-type, we trained a supervised acoustic classifier adapted from the cat call-type classification model presented by Anonymous (2027). The classifer uses a frozen ATST-Cat acoustic encoder to extract frame-level representations, followed by a single-layer bidirectional GRU and a linear classifier. The model uses independent per-class outputs to allow multiple call types to occur within the same vocalization. It was trained on synthetic multi-call clip created from a dataset of 4,502 labeled cat calls, including both sequential and overlapping vocalizations, with audio augmentation applied during training. The classifier achieved a macro F1 of 79.0% for single-call clips and 74.0% for multi-call clips (Anonymous, 2027).

The model classifies cat vocalizations into eight broad types: *chatter*, *growl*, *hiss/spit*, *meow*, *purr*, *pain cry*, *trill*, and *yowl/howl*. Manual inspection revealed reduced performance for the relatively noisy categories, particularly *hiss/spit*, *growl*, and *trill*. Additionally, call-types such as *yowl/howl* and *chatter* were frequently misclassified as *meows*. We retrained the model using a curated training set consisting of **4,250 MUs**. These MUs were manually reviewed by two expert annotators, who listened to each sample and agreed upon its final call-type label. We retained the eight original categories and introduced a ninth call-type category, *trill-meow*, to account for the frequency of this unique vocalization type (Schötz, 2012). A list of MU class definitions and counts can be found in Appendix D.

As audio extracted from social media contains significant amounts of background noise and differing recording quality, we introduced data augmentation techniques (time shifting, gain perturbation, and additive Gaussian noise) in this step to improve the robustness of the classification model (MacIsaac et al., 2024). The model achieved a macro-F1 of **79.6%** and an overall accuracy of **78.4%** on the hold-out test set. Table 2 displays the resulting per-class performance metrics of the classification model. The resulting confusion matrix can be found in Appendix E.

2.3 Clustering of Elemental Units

2.3.1 Feature Extraction and Clustering

In addition to the call-type classification of MUs, we analyzed the finer-grained acoustic patterns within each cat vocalizations. To extract these, we experimented with two feature extraction methods, Mel-frequency Cepstral Coefficients (MFCCs) (Davis & Mermelstein, 1980) and Extended Geneva Minimalistic Standard Parameter Set (eGeMAPS) (Eyben et al., 2015), and their parameters, as well as embeddings with AVES-bio configuration (Hagiwara, 2023). The features extracted through these three methods were used as input to a Gaussian Mixture Model (GMM) (Bishop & Nasrabadi, 2006), and their resulting clusters were compared before finalizing a feature extraction method. GMMs were chosen for this task based on their established performance with clustering the

Table 2: Per-class classification performance for the cat call-type classification model on the hold-out test set ($N = 850$).

Class	**P (%)**	**R (%)**	**F1 (%)**	**Support**
Chatter/Chirp	75.0	96.4	84.4	56
Growl	88.5	95.8	92.0	49
Hiss/Spit	88.9	76.2	82.1	43
Meow	78.1	72.8	75.4	205
Purr	89.1	100.0	94.2	113
Pain-Cry	93.5	82.7	87.8	105
Trill	75.6	57.6	65.4	118
Trill-Meow	55.6	67.3	60.9	103
Yowl/Howl	74.2	79.3	76.7	58
Accuracy			**78.4%**	850
F1 Macro			79.6 %	850
F1 Micro			78.4 %	850

vocal repertoire of mammals (Liang et al., 2025; Roch et al., 2007; Peso Parada & Cardenal-López, 2014; Clink & Klinck, 2021). We found that eGeMAPS were less compatible with cat vocalizations, as they were originally designed for human frequency ranges. They performed worse on cat vocalizations that fell outside of this range, such as lower purrs. AVES performed adequately for this task, but was as robust to the noise in our data as MFCCs were. For our purposes, MFCCs had the best performance, creating separate clusters for background noise which were then disregarded in our analysis, and having the most distinct vocalization clusters overall.

In our finalized pipeline, we extracted the first 13 MFCCs, as well as their first and second derivatives, for a total of 39 coefficients. With a sampling rate of 16kHz and window and hop sizes of 25ms and 10ms respectively, this produced 25ms frames within each MU, which we are calling Elemental Cat Meow Units (ECMUs). They were scaled using a Standard Scaler and input into a GMM. We calculated Bayesian Information Criterion (BIC) and Davies–Bouldin Index (DBI) (Davies & Bouldin, 1979) scores (lower is better for both) from 4-20, which can be seen in Figures 1 and 2. They reached minimums at $k = 19$ and $k = 18$ respectively, with BIC and DBI scores of 350,070,912 and 3.3419. In order to reduce the risk of combining two distinct clusters, we made the conservative decision to move forward with $k = 19$.

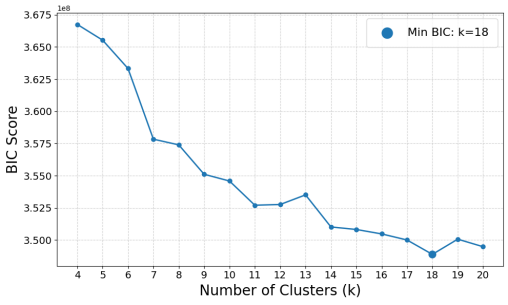

Figure 1: GMM k selection using BIC.

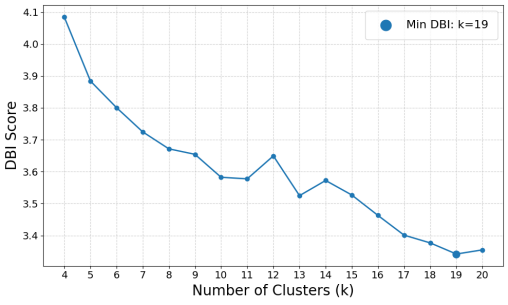

Figure 2: GMM k selection using DBI.

2.3.2 Feline and Non-feline Cluster Tests

In order to visualize the clusters output by the GMM, we projected high-dimensional embeddings using Uniform Manifold Approximation and Projection (UMAP) (McInnes et al., 2018) as seen in Figure 3. This revealed a prominent contiguous manifold (Sainburg et al., 2020) as well as a couple of smaller, distinct clusters. We hypothesized that these distant clusters represented non-feline noise,

so we chose to implement preliminary check for feline and non-feline clusters. From each cluster, we extracted 50 total samples, 10 with the highest and lowest confidences for belonging to that cluster, and 30 randomly sampled from the remaining cluster's samples. Using this method, we were able to gain an understanding of the range of sounds grouped in each cluster. To provide testers with sufficient audio to evaluate, we only played ECMUs with a continuous sequence of at least five 25ms frames of the same ECMU cluster type. After listening to these samples, we excluded 4 of the 19 original clusters from subsequent steps, as they contained largely non-feline sounds.

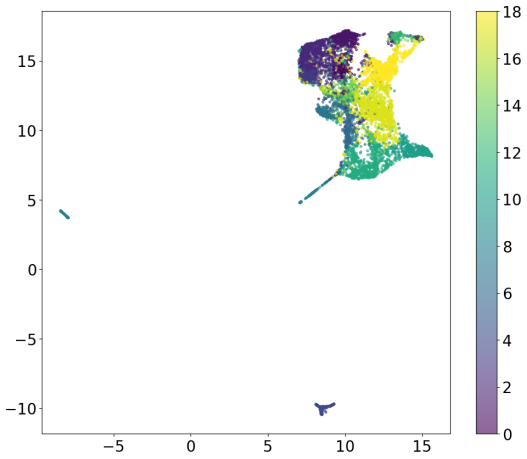

Figure 3: UMAP projection of the original 19 clusters.

2.3.3 Formal Validation of ECMU Clusters

Following the exclusion of non-feline clusters, we performed formal, double-blind ABX testing. From the remaining 15 clusters, we selected 50 random pairs of clusters to be evaluated against one another. For each set of cluster pairs, we extracted one random audio file from each cluster, assigned audio A or audio B randomly, and a third random file, audio X, from one of the two clusters being evaluated. Similar to our tests in Section 2.3.2, we only played ECMUs with continuous sequences of at least five 25ms frames with the same corresponding ECMU cluster type. Evaluators were asked to determine whether audio X belonged to cluster A or B. This was repeated five times per cluster pair for a total of 250 judgments.

We had three human experts perform these judgments, as well as a computer using soft-DTW. Human evaluators achieved an overall average accuracy of 77.3%, with 77.2% overall inter-evaluator agreement, while the computer achieved an accuracy of 68.0%. We had three cluster pairs with pairwise accuracy of 40% or less, and we revisited the extracted 50 random samples discussed in Section 2.3.2. After reviewing these cluster pairs, we were able to manually validate that these cluster pairs were indeed similar enough to warrant merging these three pairs. This action increased human accuracy to **80.0%** and computer accuracy to **68.5%** and left us with **12 final distinct clusters** containing domestic cat vocalizations.

3 Results and Analysis of Vocalization Patterns

Our analysis of domestic cat vocalization development relies on two representations of cat-call events: (i) Meow Units (MUs), and (ii) Elemental Cat Meow Units (ECMUs). We analyze these two representations in two different aspects: (i) across age, and (ii) across breeds.

3.1 Cat Call-Type Analysis

The usage of cat vocalization types across the six life stages are shown in Figure 4. Preliminary analysis of these bar charts highlights specific patterns of vocalization usage across the life stages. Statistical analysis methods are detailed in Appendix F. Figure 4 indicates a shift in dominant vo-

calization type through development, *purr* to *meow*. To test the statistical significance of this shift, we used mixed-effects logistic regression models. The results showed a significant decrease in purr usage over life stages ($\beta = -0.30$, $SE = 0.08$, $p < 0.001$) with approximately 26% lower odds with each life stage ($OR = 0.74$, $95\%CI = 0.64, 0.86$). On the other hand, *meow* usage significantly increased across life stages ($\beta = 0.14$, $SE = 0.06$, $p = 0.017$) with approximately 15% higher odds with each life stage ($OR = 1.15$, $95\%CI = 1.03, 1.29$). This shift seems to be constant across all breeds as no significant interaction was found between breed and age in these models. Another age-related trend represented in Figure 4 was an increase in *snarl/pain-cry* usage across life stages. This developmental change was tested with a chi-square test, which found that the increase was statistically significant ($\chi^2(df = 1, N = 9,135) = 56.60$, $p < 0.001$, $\Phi = 0.08$), with the odds of a *snarl/pain-cry* being used increasing by approximately 7.8 times ($OR = 7.79$, $95\%CI = 4.22, 14.39$) in older cats (mature, senior, geriatric) versus younger cats (kitten, junior, prime). This increase may be related to elevated confusion and distress seen in older cats as their health deteriorates (Černá et al., 2020; Eisinger & Kuhne, 2025), however health condition was not directly measured in this study.

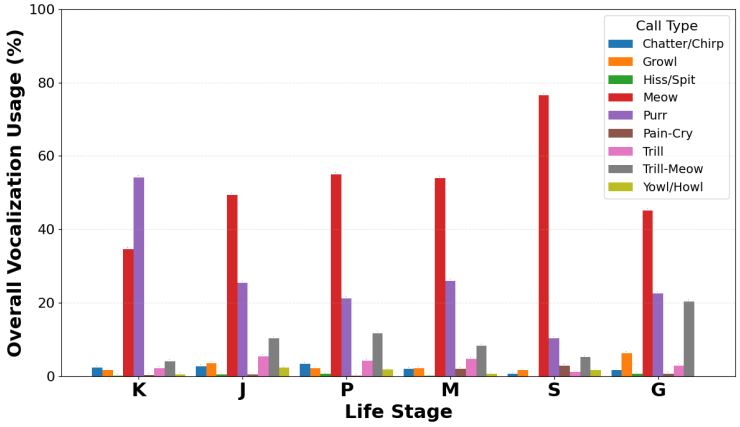

Figure 4: Bar plot showing percentage of cat call-type usage per age group averaged over all breeds.

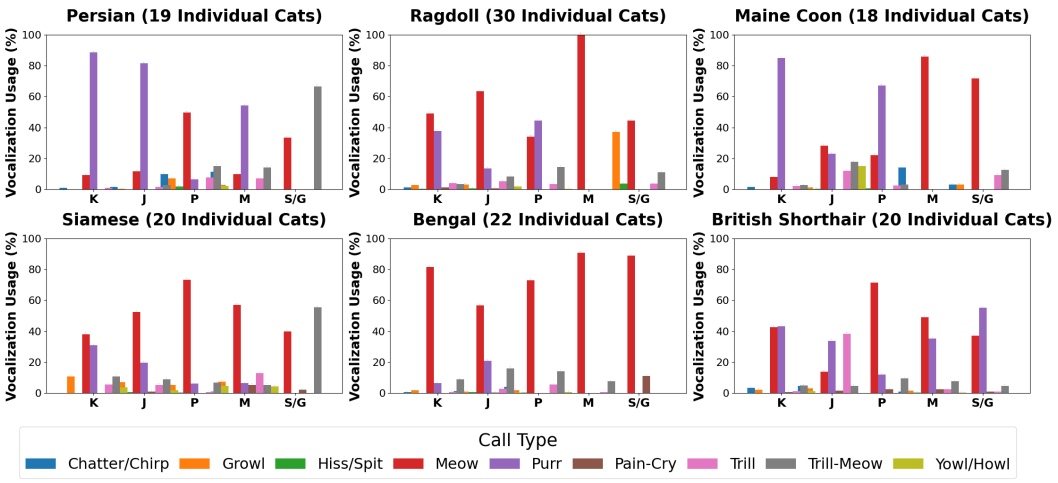

Figure 5: Bar plots showing the percentage of cat call-type usage per age group for each main breed.

In Figure 5, we group together the senior and geriatric age groups for clearer analysis of feline vocalization across age due to scarcer data for older cats. Figure 5 shows cat vocalization usage across life stages for each of the main six breeds, highlighting unique breed variations. In Figure 5, the distribution of vocalization types across life stages differs between the breeds and Shannon

entropy was used to compare these differences. Ragdoll ($H = 0.935$) and Siamese ($H = 0.916$) cats had the highest mean Shannon indexes of the main six breeds, indicating a more even distribution of vocalization types across ages. Meanwhile, Persian cats had a much lower mean Shannon entropy ($H = 0.656$), highlighting the domination of fewer vocalization types across life stages. However, each breed had differing amounts of variation between individuals, with Persian cats having the most intraspecies variation ($SD = 0.60$) and British Shorthair cats having the least ($SD = 0.33$).

3.2 Elemental Unit Cluster Analysis

Our audio processing from Section 2.3 resulted in 12 clusters of ECMU types across all breeds and life stages, shown in Figure 6. Upon preliminary analysis, we found that the variety and frequency of the ECMUs were relatively similar across all cat breed types, with a few exceptions.

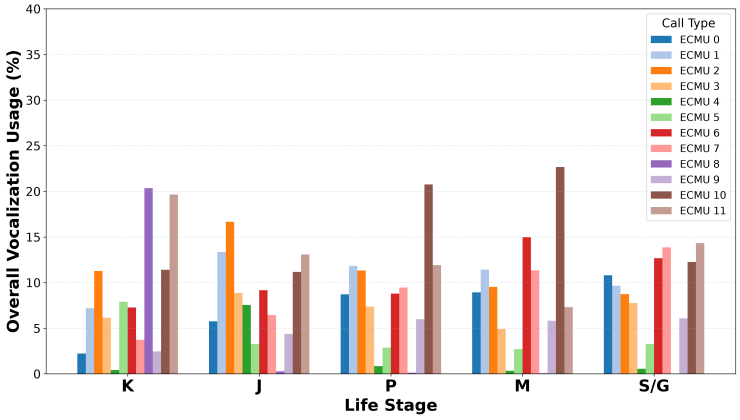

Figure 6: Bar plot showing the averaged ECMU cluster usage per age group across all cats.

ECMUs 0 and 7 consistently increase in usage throughout the lifespans of the cats. Beta-binomial mixed-effects models indicated that ECMU 0 increased from approximately 4% in kittens to 8% in S/G cats across life stages ($p = 0.045$). Similarly, ECMU 7 increased from approximately 6.5% in kittens to 11% in S/G cats ($p = 0.043$). These are likely part of calls used in a more complicated context or interaction, as older cats tend to use them more. The most frequent type is ECMU 10, having over 10% ECMU type usage over every life stage. This indicates that ECMU 10 is versatile and can be used in many contexts and interactions. Interestingly, every cat breed had a similar number of ECMU types (12). This is in contrast to the findings of Lekhak et al. (2025), which found significant differences between EDBU (Elemental Dog Bark Units) type number, diversity, and usage per life stage across dog breeds.

4 Conclusion

We present a rich dataset that opens avenues for future longitudinal studies on animal communication, along with unique findings on age-related changes in the vocal repertoire of cats. Future research should aim to expand the dataset in order to strengthen underrepresented breeds and age groups, as this will broaden the overall applicability and improve statistical relevance. This dataset can be further deepened through comparing the finer grain parameters of each ECMU between and within breeds through the creation of an acoustic ‘guide’. Future work could also explore the situational meanings behind MUs and ECMUs by investigating the correlating contexts or semantics involved with these units. Connecting age-related changes in cat vocalizations with other known developmental changes allows for a more complete view of aging in cats, which is important for veterinary and pet care. Additionally, this dataset presents an opportunity for further exploration into advanced representation learning methods to capture feline bioacoustics, improving upon the results from acoustic features such as MFCCs or eGeMAPS. Through these new avenues, this dataset allows for significant advancements in computational animal linguistics, and expands knowledge of changes in feline vocal communication across their lifespans.

5 AI USE STATEMENT

We used generative AI tools, including Google's Gemini (Team et al., 2024), and Anthropic's Claude (Anthropic, 2024) to help write and debug code designed by the authors related to clustering and batch file processing. ChatGPT-5.6 Luna (OpenAI, 2026) was used to help generate and draft initial code for statistical analysis. The authors reviewed and edited all AI-assisted work to verify function and correctness, and we take full responsibility for the final code, results, interpretation of results, and content of the research presented.

REFERENCES

Anonymous. Catnet: Discovering recurring domestic cat vocalizations from web video, 2027. Submitted to AAAI 2027. Under review.

Anthropic. Claude 3.5 sonnet, 2024. URL `https://www.anthropic.com/news/claude-3-5-sonnet`. Accessed: 2026.

Douglas Bates, Martin Mächler, Ben Bolker, and Steve Walker. Fitting linear mixed-effects models using lme4. *Journal of statistical software*, 67:1–48, 2015.

Jan Bellows, Sharon Center, Leighann Daristotle, Amara H Estrada, Elizabeth A Flickinger, Debra F Horwitz, B Duncan X Lascelles, Allan Lepine, Sally Perea, Margie Scherk, et al. Aging in cats: common physical and functional changes. *Journal of Feline medicine and surgery*, 18(7):533–550, 2016.

Christopher M Bishop and Nasser M Nasrabadi. *Pattern recognition and machine learning*, volume 4. Springer, 2006.

Petra Černá, Hannah Gardiner, Lorena Sordo, Camilla Tørnqvist-Johnsen, and Danièlle A Gunn-Moore. Potential causes of increased vocalisation in elderly cats with cognitive dysfunction syndrome as assessed by their owners. *Animals*, 10(6):1092, 2020.

Dena J Clink and Holger Klinck. Unsupervised acoustic classification of individual gibbon females and the implications for passive acoustic monitoring. *Methods in Ecology and Evolution*, 12(2): 328–341, 2021.

David L Davies and Donald W Bouldin. A cluster separation measure. *IEEE transactions on pattern analysis and machine intelligence*, (2):224–227, 1979.

Steven Davis and Paul Mermelstein. Comparison of parametric representations for monosyllabic word recognition in continuously spoken sentences. *IEEE transactions on acoustics, speech, and signal processing*, 28(4):357–366, 1980.

Sagi Denenberg, Karen L Machin, and Gary M Landsberg. Behavior and cognition of the senior cat and its interaction with physical disease. *Veterinary Clinics: Small Animal Practice*, 54(1): 153–168, 2024.

Joana Eisinger and Franziska Kuhne. The impact of disease on behavior: Altering behavior in the course of disease in aging cats. In *Pets*, volume 2, pp. 21. MDPI, 2025.

Florian Eyben, Klaus R Scherer, Björn W Schuller, Johan Sundberg, Elisabeth André, Carlos Busso, Laurence Y Devillers, Julien Epps, Petri Laukka, Shrikanth S Narayanan, et al. The geneva minimalistic acoustic parameter set (gemaps) for voice research and affective computing. *IEEE transactions on affective computing*, 7(2):190–202, 2015.

Adam R Fishbein, Jonathan B Fritz, William J Idsardi, and Gerald S Wilkinson. What can animal communication teach us about human language? *Philosophical Transactions of the Royal Society B: Biological Sciences*, 375(1789):20190042, 2019.

Jort F Gemmeke, Daniel PW Ellis, Dylan Freedman, Aren Jansen, Wade Lawrence, R Channing Moore, Manoj Plakal, and Marvin Ritter. Audio set: An ontology and human-labeled dataset for audio events. In *2017 IEEE international conference on acoustics, speech and signal processing (ICASSP)*, pp. 776–780. IEEE, 2017.

Governing Council of the Cat Fancy. 2025 registered breeds by ranking. https://gccfcats.org, 2026. URL https://www.gccfcats.org/wp-content/uploads/2026/01/Registered-Cats-By-Breed-2024-2025.pdf. Accessed: 2026-07-17.

Aaron Grattafiori, Abhimanyu Dubey, Abhinav Jauhri, Abhinav Pandey, Abhishek Kadian, Ahmad Al-Dahle, Aiesha Letman, Akhil Mathur, Alan Schelten, Alex Vaughan, et al. The llama 3 herd of models. *arXiv preprint arXiv:2407.21783*, 2024.

Jasmine MS Grimsley, Jessica JM Monaghan, and Jeffrey J Wenstrup. Development of social vocalizations in mice. *PloS one*, 6(3):e17460, 2011. doi: https://doi.org/10.1371/journal.pone.0017460.

Yasemin B Gultekin, David GC Hildebrand, Kurt Hammerschmidt, and Steffen R Hage. High plasticity in marmoset monkey vocal development from infancy to adulthood. *Science Advances*, 7(27):eabf2938, 2021.

Masato Hagiwara. Aves: Animal vocalization encoder based on self-supervision. In *ICASSP 2023-2023 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pp. 1–5. IEEE, 2023.

Roland Kays, Robert R Dunn, A Waldstein Parsons, Brandon McDonald, Troi Perkins, SA Powers, Leonora Shell, Jenni L McDonald, Holly Cole, Heidy Kikillus, et al. The small home ranges and large local ecological impacts of pet cats. *Animal Conservation*, 23(5):516–523, 2020.

Raymond D Kent and Ann D Murray. Acoustic features of infant vocalic utterances at 3, 6, and 9 months. *The Journal of the Acoustical Society of America*, 72(2):353–365, 1982.

Qiuqiang Kong, Yin Cao, Turab Iqbal, Yuxuan Wang, Wenwu Wang, and Mark D Plumbley. Panns: Large-scale pretrained audio neural networks for audio pattern recognition. *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, 28:2880–2894, 2020.

Hridayesh Lekhak, Tuan M Dang, Theron S Wang, and Kenny Q Zhu. A data-driven approach to the longitudinal study of canine vocal pattern development. In *Proceedings of the 33rd ACM International Conference on Multimedia*, pp. 12473–12482, 2025.

Yinian Liang, Yan Wang, Fangjiong Chen, Hua Yu, Fei Ji, and Yankun Chen. Automatic detection and unsupervised clustering-based classification of cetacean vocal signals. *Applied Sciences*, 15 (7):3585, 2025.

Monika J Lipinski, Lutz Froenicke, Kathleen C Baysac, Nicholas C Billings, Christian M Leutenegger, Alon M Levy, Maria Longeri, Tirri Niini, Haydar Ozpinar, Margaret R Slater, et al. The ascent of cat breeds: genetic evaluations of breeds and worldwide random-bred populations. *Genomics*, 91(1):12–21, 2008.

Xubo Liu, Haohe Liu, Qiuqiang Kong, Xinhao Mei, Jinzheng Zhao, Qiushi Huang, Mark D Plumbley, and Wenwu Wang. Separate what you describe: Language-queried audio source separation. *arXiv preprint arXiv:2203.15147*, 2022.

Luca A Ludovico, Stavros Ntalampiras, Giorgio Presti, Simona Cannas, Monica Battini, and Silvana Mattiello. Catmeows: A publicly-available dataset of cat vocalizations. In *International Conference on Multimedia Modeling*, pp. 230–243. Springer, 2021.

L Lyons. Cat domestication and breed development. In *10th World Congress on Genetics Applied to Livestock Production. Canada*, 2014.

Jennifer MacIsaac, Stuart Newson, Adham Ashton-Butt, Huma Pearce, and Ben Milner. Improving acoustic species identification using data augmentation within a deep learning framework. *Ecological Informatics*, 83:102851, 2024.

Leland McInnes, John Healy, and James Melville. Umap: Uniform manifold approximation and projection for dimension reduction. *arXiv preprint arXiv:1802.03426*, 2018.

OpenAI. Chatgpt (gpt-5.6 luna version), 2026. URL https://chat.openai.com. Accessed: 2026.

Pablo Peso Parada and Antonio Cardenal-López. Using gaussian mixture models to detect and classify dolphin whistles and pulses. *The Journal of the Acoustical Society of America*, 135(6): 3371–3380, 2014.

Jessica Quimby, Shannon Gowland, Hazel C Carney, Theresa DePorter, Paula Plummer, and Jodi Westropp. 2021 aaha/aafp feline life stage guidelines. *Journal of Feline Medicine and Surgery*, 23(3):211–233, 2021.

R Core Team. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria, 2025. URL `https://www.R-project.org/`.

Alec Radford, Jong Wook Kim, Chris Hallacy, Aditya Ramesh, Gabriel Goh, Sandhini Agarwal, Girish Sastry, Amanda Askell, Pamela Mishkin, Jack Clark, et al. Learning transferable visual models from natural language supervision. In *International conference on machine learning*, pp. 8748–8763. PmLR, 2021.

Md Saifur Rahman and Hassan Reza. A systematic review towards big data analytics in social media. *Big data mining and analytics*, 5(3):228–244, 2022.

Marie A Roch, Melissa S Soldevilla, Jessica C Burtenshaw, E Elizabeth Henderson, and John A Hildebrand. Gaussian mixture model classification of odontocetes in the southern california bight and the gulf of california. *The Journal of the Acoustical Society of America*, 121(3):1737–1748, 2007.

Tim Sainburg, Marvin Thielk, and Timothy Q Gentner. Finding, visualizing, and quantifying latent structure across diverse animal vocal repertoires. *PLoS computational biology*, 16(10):e1008228, 2020.

Susanne Schötz. A phonetic pilot study of vocalisations in three cats. In *FONETIK 2012*, pp. 45–48. University of Gothenburg, 2012.

Susanne Schötz, Joost van de Weijer, and Robert Eklund. Context effects on duration, fundamental frequency, and intonation in human-directed domestic cat meows. *Applied Animal Behaviour Science*, 270:106146, 2024.

Jerzy P Szaflarski, Vincent J Schmithorst, Mekibib Altaye, Anna W Byars, Jennifer Ret, Elena Plante, and Scott K Holland. A longitudinal functional magnetic resonance imaging study of language development in children 5 to 11 years old. *Annals of Neurology: Official Journal of the American Neurological Association and the Child Neurology Society*, 59(5):796–807, 2006.

Lu Tang, Shipeng Hu, Choujun Yang, Rui Deng, Aibin Chen, and Guoxiong Zhou. Jl-tfmsfnet: A domestic cat sound emotion recognition method based on jointly learning the time–frequency domain and multi-scale features. *Expert Systems with Applications*, 255:124620, 2024.

Chloé Tavernier, Sohail Ahmed, Katherine Albro Houpt, and Seong Chan Yeon. Feline vocal communication. *Journal of veterinary science*, 21(1):e18–e18, 2020.

Gemini Team, Petko Georgiev, Ving Ian Lei, Ryan Burnell, Libin Bai, Anmol Gulati, Garrett Tanzer, Damien Vincent, Zhufeng Pan, Shibo Wang, et al. Gemini 1.5: Unlocking multimodal understanding across millions of tokens of context. *arXiv preprint arXiv:2403.05530*, 2024.

Astrid van Toor, Nadeem Qazi, and Stefania Paladini. A deep learning pipeline for age prediction from vocalisations of the domestic feline. *Scientific Reports*, 15(1):34565, 2025.

Amy Hoyumpa Vogt, Ilona Rodan, Marcus Brown, Scott Brown, CA Tony Buffington, MJ Larue Forman, Jacqui Neilson, and Andrew Sparkes. Aafp–aaha: Feline life stage guidelines. *Journal of the American Animal Hospital Association*, 46(1):70–85, 2010.

Hadley Wickham, Romain François, Lionel Henry, Kirill Müller, and Davis Vaughan. *dplyr: A Grammar of Data Manipulation*, 2026. URL `https://dplyr.tidyverse.org`. R package version 1.2.1.

World Population Review. Youtube users by country 2026. `https://worldpopulationreview.com/country-rankings/youtube-users-by-country`, 2026. Accessed: 2026-07-17.

A AUDIOSEP TESTING

Table 3: AudioSep manual performance evaluation on a random test set ($N = 195$).

Rating	**# Videos** (Rater 1 / 2)	**% Videos** (Rater 1 / 2)
5	13 / 14	6.67% / 7.18%
4	120 / 119	61.54% / 61.03%
3	44 / 45	22.56% / 23.08%
2	17 / 16	8.72% / 8.21%
1	1 / 1	0.51% / 0.51%

B BINARY FELINE-VOCALIZATION CLASSIFICATION PERFORMANCE

Table 4: Per-class classification performance for the binary feline-vocalization classification model on the hold-out test set ($N = 963$).

Class	**P (%)**	**R (%)**	**F1 (%)**	**Support**
Cat Vocalization	99.5	98.4	99.0	870
Background Noise	86.5	95.7	90.9	93
Accuracy			**98.1%**	963

C DATASET OVERVIEW

Table 5: Final dataset statistics after Meow Units segmentation and filtering (main 6 breeds).

Metric	Bengal	BSH	Maine Coon	Persian	Ragdoll	Siamese
Individuals	21	15	16	16	27	18
Videos	1,205	532	260	646	777	581
MU Count	2,254	1,000	716	626	1,922	1,366
MU total dur (h)	1.13	0.48	0.32	0.25	0.96	0.63

Table 6: Age groups: No. of unique cats (total MUs) (main 6 breeds, $N = 129$).

Metric	Bengal	BSH	Maine Coon	Persian	Ragdoll	Siamese
Kitten	16 (160)	12 (120)	11 (301)	10 (100)	18 (271)	13 (73)
Junior	21 (773)	10 (65)	14 (244)	11 (251)	27(1140)	15 (849)
Prime	4 (1168)	7 (87)	8 (129)	10 (190)	10 (476)	11 (277)
Mature	1 (135)	5 (580)	3 (7)	3 (82)	1 (2)	4 (110)
Senior	1 (18)	4 (47)	0	2 (1)	0	0
Geriatric	0	1 (61)	1 (35)	2 (2)	1 (33)	1 (57)

Table 7: Final dataset statistics after Meow Units segmentation and filtering (supplemental 6 breeds).

Metric	Aby	Birman	Devon Rex	Siberian	Sphynx	Tonkinese
Individuals	3	3	4	6	2	4
Videos	119	189	60	203	93	23
MU Count	210	311	138	465	106	21
MU total dur (h)	0.12	0.16	0.06	0.24	0.04	0.01

Table 8: Age groups: No. of unique cats (total MUs) (supplemental 6 breeds, $N = 27$).

Metric	Aby	Birman	Devon Rex	Siberian	Sphynx	Tonkinese
Kitten	2(86)	3 (1)	3 (45)	1 (1)	2 (2)	3 (2)
Junior	3 (120)	1 (27)	1 (79)	3 (19)	0	3 (13)
Prime	2 (2)	1 (283)	1 (3)	4 (379)	0	2 (6)
Mature	0	0	0	2 (52)	0	0
Senior	1 (2)	0	0	2 (14)	1 (104)	0
Geriatric	0	0	1 (1)	0	0	0

D CAT CALL-TYPE CLASS DEFINITIONS

Table 9: Cat call-type class definitions and count in training set ($N = 4,250$).

Class	**Definition**	**MU Count**
Meow	The distinctive, canonical vocalization of cats	1,023
Trill	Greeting call produced with a soft voice and a rising inflection, similar to a purr	589
Trill-Meow	Hybrid of the meow and trill, beginning with a closed mouth and ending with an open-mouthed meow	514
Purr	Low-frequency, continuous murmur sound produced on the inhale and exhale with a closed mouth	565
Howl/Yowl	Harsh, drawn-out vocalizations with variable pitch, tone, and intensity	289
Growl	A low-pitched, sustained rumbling noise in the throat associated with aggression	247
Pain Cry	Intense, strained cries associated with distress or aggression	527
Chatter/Chirp	Rhythmic clicking of the jaw or staccato sounds / short, high-pitched sounds (both used to imitate birds or rodents)	280
Hiss/Spit	Rapid, forceful expulsion of air, often with a hard onset	216

E CAT CALL-TYPE CLASSIFICATION PERFORMANCE

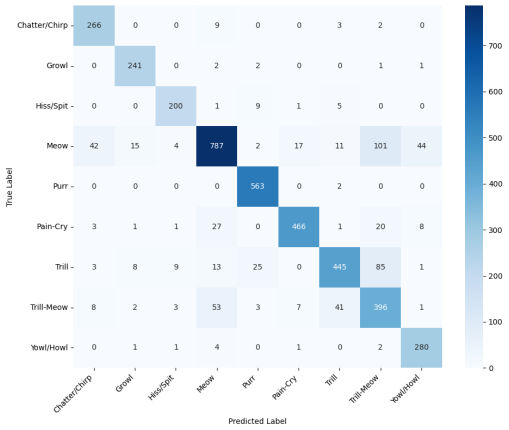

Figure 7: Confusion Matrix for Cat Call-Type Classifier by Counts.



Figure 8: Confusion Matrix for Cat Call-Type Classifier by Percentage.

F STATISTICAL ANALYSIS

All statistical calculations for the cat call-type analysis results were done using R version 4.5.2 (R Core Team, 2025). To test how *purr* and *meow* usage changed over life stages, we used Binomial generalized linear mixed-effects logistic regression models with a logit link. This was performed with *purr* as the binary dependent variable (1 = yes, 0 = no), and again with *meow* as the dependent variable. Life stage was ordered numerically from 1 (kitten) to 6 (geriatric) and was treated as a fixed effect, while cat ID was treated as a random effect due to the dataset including multiple vocalization events from the same individual cats. These tests were then repeated, adding in breed as a fixed effect in order to see if there was an interaction between breed and age for *purr* and *meow* usage. The results were analyzed with significance level ($\alpha = 0.05$). The models were fitted using maximum liklihood with the glmer function in the `lme4` package (version 2.0.6; Bates et al., 2015). To test if the relationship between age and *snarl/pain-cry* usage was statistically significant, a chi-square test of independence with Yates' continuity correction was used as all cell counts were greater than 5. The null hypothesis for this test was that there is no statistically significant relationship between age and *snarl/pain-cry* usage. The results were also analyzed with significance level ($\alpha = 0.05$). Finally, to quantify the variation in distribution of vocalization types across life stages between the breeds, Shannon entropy was calculated for each individual cat and then organized by breed. The following equation was used for this calculation:

$$H(X) = -\sum_{i=1}^{n} p(x_i) \ln p(x_i), \tag{1}$$

where x_i was the distinct vocalization type, and $p(x_i)$ was estimated as number of vocalization type x_i divided by the total number of vocalizations for that cat. Shannon entropy was calculated using natural logarithms. Zero-handling was not needed as vocalization types with zero observations for an individual were not included in the calculation. The `dplyr` package (version 1.2.1; Wickham et al., 2026) was used in these calculations.

Statistical calculations for ECMU analysis were calculated in a similar fashion, using beta-binomial mixed effects modeling through R. ECMU type, life stage, and breed were fixed effects and cat ID was a random effect. Like the cat-call statistical calculations, the models were fitted using maximum liklihood with the glmer function in the `lme4` package (Bates et al., 2015). The life stages Senior and Geriatric were combined to balance out the dataset, as some breeds did not have data for those life stages. The results were analyzed with significance level ($\alpha = 0.05$).