scientific reports

Check for updates

OPEN

A deep neural network for automatic prediction of mouse behavior from ultrasonic vocalizations emitted during social interaction test

Alessandro Gnutti[1,3]✉, Marika Premoli[2,3], Chia-Hao Kao[1], Pierangelo Migliorati[1], Riccardo Leonardi[1] & Sara Anna Bonini[2]

The study of ultrasonic vocalizations (USVs) is fundamental as they represent a primary form of communication in mice, offering valuable insights into their social behaviors and underlying neural processes. This study introduces the first deep learning system designed to predict mouse behavior using only USVs as input. Applied to both wild-type (WT) and p50 knockout (KO) transgenic mice that present behavioral deficits, our results reveal a clear correlation between mouse communication and behavior. We also observed asymmetric cross-genotype generalization: models trained on WT USVs generalize to KO mice, whereas models trained on KO data fail to generalize to WT, suggesting differences in the distribution of spectrotemporal patterns between genotypes.

Keywords Ultrasonic vocalizations, Mouse behavior, Mouse communication, Correlation study, Automatic prediction, Deep-neural-network classifier

Ultrasonic vocalizations (USVs) are emitted by mice for communication between themselves, typically ranging from 30 kHz to 110 kHz[1,2]. These vocalizations play a significant role in mouse interactions across different life stages[3]. As infants, pups vocalize to attract maternal care[4], while in adolescence and adulthood, mice produce USVs during social interactions which are associated with social bonding[5,6] as well as social and sexual behaviors[7]. Despite the importance of these vocalizations, the relationship between USVs and specific behaviors remains largely unexplored. Some studies have attempted to establish this link. For example, Asaba and colleagues[8] investigated the function of the USVs emitted during male-female interaction, demonstrating USVs importance in sexual attractiveness and reproductive success. However, most studies have focused on more simplistic handcrafted features, such as the patterns and shapes of USVs[9–11], thereby limiting the scope of exploration and findings. In particular, the study in[9] reported distinct USV patterns associated with specific social activities performed by male mice. In that work, behavioral decoding was based on manually defined acoustic descriptors and vocal categories, followed by statistical classification. Similarly, de Chaumont et al.[10] combined long-term behavioral tracking with call-level acoustic analysis to investigate the relationship between vocalizations and social interactions. These studies have contributed to clarifying how specific vocal patterns relate to behavioral states, primarily through segmentation-based and feature-level characterization. Additionally, our research group examined the diversity of ultrasonic communication patterns emitted by two groups of mice, identifying significant differences in pattern lengths between the groups[11].

More recently,[12] explored the link between USVs emitted by mouse pups and their behavior during adolescence. They found a strong positive correlation between call rate and sociability: mice that were high vocalizers as pups exhibited more prosocial behavior later in life. According to their findings, the number of neonatal USVs can serve as a predictor of future prosocial tendencies in mice. In parallel, recent methodological advances have also improved the precision of USV attribution during social interactions. For example, Sterling et al.[13] introduced a high-precision hybrid acoustic localization system to assign vocalizations to individual emitters, while Oliveira-Stahl et al.[14] developed refined localization algorithms and analyzed how acoustic

[1]Department of Information Engineering, University of Brescia, Brescia, Italy. [2]Department of Molecular and Translational Medicine, University of Brescia, Brescia, Italy. [3]Alessandro Gnutti and Marika Premoli contributed equally to this work. ✉email: alessandro.gnutti@unibs.it

features vary with spatial relationships, sex, and genotype. These studies significantly enhance the spatial and acoustic characterization of mouse communication. However, they do not address the automatic prediction of behavior directly from vocal signals.

In the last few years, artificial intelligence has gained significant importance in this field and beyond. While several studies have employed deep learning techniques, primarily for USV call segmentation, detection and classification[15–21], none has investigated behavior prediction starting from USVs as input. In this paper, we present the first deep-learning-based algorithm capable of automatically recognizing mouse behavior on the base of USV communication. Unlike previous approaches based on predefined acoustic features or vocal categories, the proposed framework operates in an end-to-end manner, directly learning discriminative representations from spectrogram inputs without explicit feature engineering.

Our analysis encompasses both control wild-type (WT) mice and p50 knockout (KO) mice. p50 KO mice have a deletion of the NFKB1 gene, coding for the NF-κB p50 subunit. NF-kB is an important transcription factor, ubiquitously expressed, that regulates several functions such as inflammation, cell survival and death[22,23], in cell plasticity and morphology remodeling[24–26] and neurogenesis[27]. Our research group characterized p50 KO mice as a model of neurodevelopmental disorders, indeed p50 KO mice have an abnormal columnar organization in the somatosensory cortex and altered neurite orientation, associated with hyperactivity, social interaction deficit, altered ultrasonic communication[28–30]. All these cortical and behavioural alterations have also been reported in several NDDs models[31,32]. In this paper, we considered this model of neurodevelopmental disorders and curated a dataset by recording audio-video files of mice activities and annotating mouse behaviors through video analysis. Then, we designed a deep neural network (DNN) classifier[33,34] to identify the behavior class from the pre-processed audio clips. By comparing the prediction results with ground truth annotations, we aim to uncover correlations between USVs and behaviors (see Fig. 1 for an overview of the proposed high-level framework).

Materials

39 male wild-type (WT) (B6;129PF2) and 37 male NF-κB p50 knock-out (KO) (B6;129PF2- NF-κB 1tm 1 Bal/J) adult mice (3 months of age) were used. For the social test, 10 adult WT and 10 adult KO female mice were used as stimulus mice. Mice were purchased from the Charles River Laboratories (Calco, Italy) and housed in groups of 3-4 mice in standard cages (15 cm wide x 35 cm long x 12 cm deep) in our animal facility in a 12 hours light/dark cycle (lights off at 19:00) with food and water ad libitum. Temperature (23$\pm$ 2°C) and humidity (55±15%) in the cage were automatically regulated by the Sealsafe Aero System by individually ventilated cages with EPA filters (Tecniplast Group, Italy). All experiments were performed according to European Communities Council Directive guidelines (CEE N°86/609) and approved by the Animal Welfare Committee (Organismo Preposto al Benessere Animale, OPBA) of the University of Brescia and by the Italian Ministry of Health.

Methods

Behavioral and USV recording

To observe mice interactions, male mice were isolated in their cages for 3 days and on fourth day an unfamiliar female of the same genotype was introduced in the cage of isolated male (resident male) for 5 minutes. On the day of reciprocal social interaction, the vaginal estrous phase of female mice was checked as previously described in[35]. They were distributed in a balanced way, that is, the experimental males of the same experimental group encountered half of stimulus females that were in diestrous and half of stimulus female in estrous. A camera (Basler monochrome Near Infra-Red GigE camera, Noldus, The Netherlands), positioned 20 cm above the cage, recorded their behavior for five minutes. Mouse behaviors were categorized into two broad groups: social behaviors, which include body sniffing, anogenital sniffing, nose sniffing, mounting, following, and contact; and non-social behaviors, including cage exploring, digging, self-grooming, and rearing. This classification (social and non-social behavior) was chosen to simplify the analysis while capturing the essential differences between

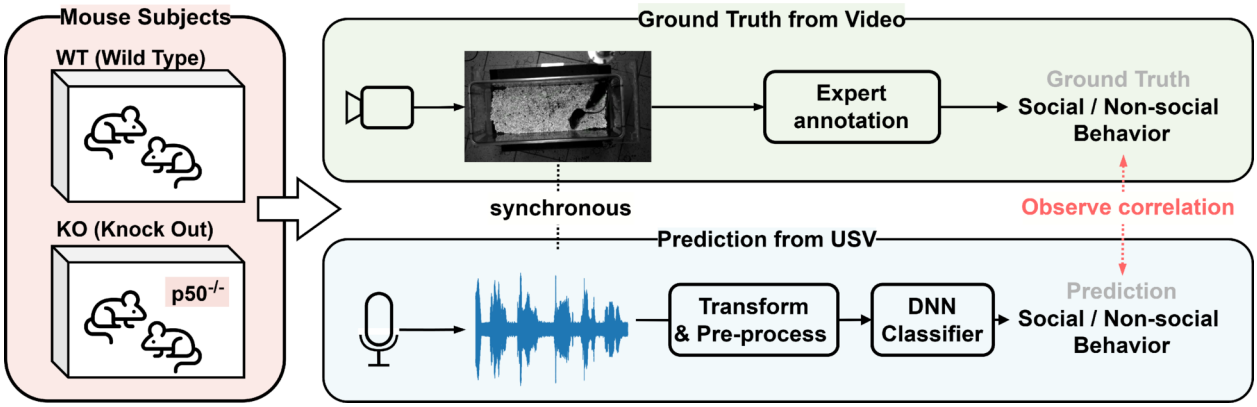

Fig. 1. High-level system overview applied to both WT and KO mice (pink box). Ground truth acquisition (green box): video recording with manually annotated behaviors. Behavior prediction (blue box): USV recording and classification.

interactions involving direct engagement with another mouse and those that do not. Specifically, the time spent by test mice participating in both social behaviors and non-social behaviors were observed and quantified by two independent operators obtaining consistent behavioral ratings. Observer XT (version 14.1, Noldus, The Netherlands), a specialized commercial software for behavioral analysis[32,36] was used. During the test, USVs emitted by the mice were recorded using a microphone, placed vertically 20 cm above the cage, with a sample rate at 250 kHz sensitive to ultrasound frequencies (CM16/CMPA, Avisoft Bioacoustics, Berlin), as previously described in[30,32,37]. In total, we collected 76 video and audio files, each 5 minutes long, consisting of 39 from WT mice and 37 from KO mice.

High-level framework

To explore the correlation between mice USVs and behaviors, we designed a Deep Neural Network (DNN) acting as a classifier for the pre-processed audio clips. The objective of this classifier is to extract features from the associated audio spectrograms and then determine whether the given clip corresponds to a social or non-social behavior class. The overall framework is illustrated in Fig. 2. The convolutional-based DNN processes the spectrogram as its input, in a way similar to an image classification task, and outputs probabilities for the two classes. The predicted behavior class is determined by the one with the highest assigned probability.

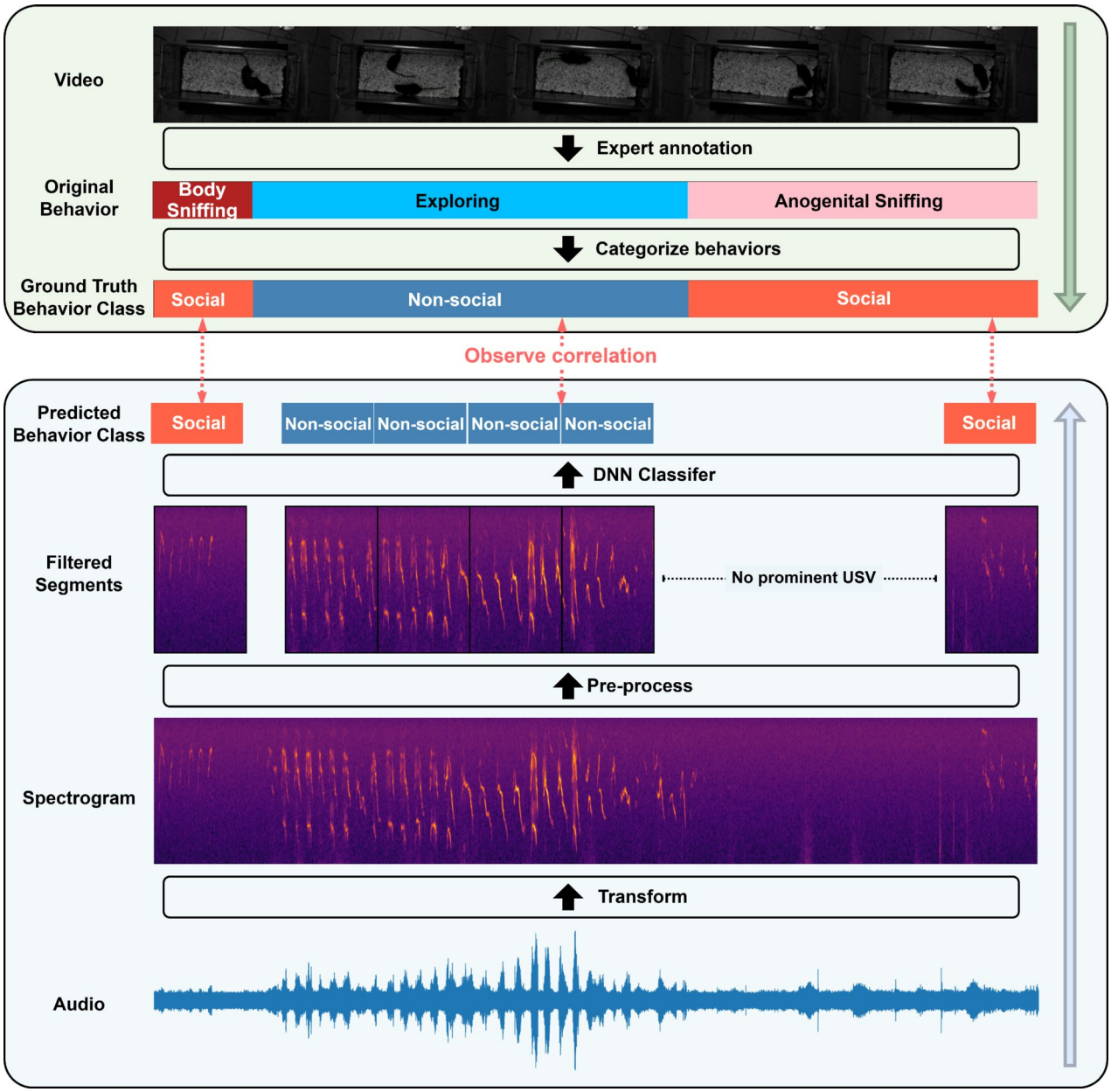

Fig. 2. Detailed pipeline of proposed system with exemplary samples.

For training the DNN classifier, the audio clips are divided into three distinct sets: one for training the neural network (70%), one for validation to optimize hyperparameters (15%), and one for testing (15%). Additionally, given the distinct nature of WT and KO mice, we trained two separate neural networks for each strain of mice and experimented with different segment durations *d*. This allows us to account for potential differences in vocalization patterns and behavioral characteristics between the two groups of mice.

Pre-processing

The captured audio clips, synchronized with the captured videos, undergo pre-processing prior to the classification by the neural network, as illustrated in Fig. 3. Given that mice emit USVs only sporadically (more than 65% of the recorded audio contains no vocal activity), the recorded audio is filtered to extract fixed-length segments of duration *d* containing a minimum proportion of ultrasonic vocalizations. Since USV signals are more distinguishable in the time-frequency domain, each segment is firstly converted into a mel-scale spectrogram. Then, to identify USVs for each time instance, the maximum magnitude is examined along the frequency axis of the spectrogram. If this value exceeds a predetermined threshold T_n, it is considered to contain USV signals in that time instance. Finally, the percentage of the segment containing USVs is computed, and only segments where this percentage surpasses a second threshold T_p are included in the dataset. Note that the minimum percentage threshold T_p is applied independently of the chosen segment duration *d*, ensuring that each retained segment contains a sufficient proportion of vocal activity regardless of its absolute length. These thresholds ensure that only segments with meaningful vocalizations are retained for the analysis. This selection criterion is aligned with the objective of the study, which is to assess whether behavior can be predicted specifically from vocal communication. Including silent or near-silent segments would introduce contextual information unrelated to acoustic signaling, therefore, only segments containing sufficient USV activity are retained for analysis.

To determine the appropriate segment duration *d*, the distributions of behavior duration, based on the human annotation, are analyzed and shown in Fig. 4A. As the behaviors are observed to have shorter durations, typically ranging from 0 to 2 seconds, we designed the experiments to include segment lengths $d = \{0.5, 1, 1.5, 2\}$ seconds. For all experiments, the detection threshold T_n is set to -10 dB, while the minimum percentage threshold T_p is fixed at 10%, independently of the selected segment duration. Figure 4B shows the resulting number of pre-processed training and testing segments for WT and KO mice across different durations. We recall that the complete audio–video recording from a given mouse was assigned exclusively to a single subset (training, validation or test) before any segmentation was performed. This subject-level separation ensures that all segments derived from the same animal belong to the same split, preventing information leakage and avoiding potential repeated-measures bias. Notably, to prevent label ambiguity during training, we retained only segments fully associated with a single annotated behavior; segments overlapping multiple behaviors or behavioral transitions were discarded.

DNN classifier

We adopted a DNN-based classifier designed to process input spectrograms. As illustrated in Fig. 5, the model comprises a feature extraction module followed by a classification head. The feature extraction module consists of three convolutional layers, each employing a 3×3 kernel with a stride of 1 and padding of 1. Each convolution is followed by Batch Normalization, a ReLU activation function, and a 2×2 Max Pooling layer with a stride of 2 to down-sample the spatial resolution by half. The number of channels progressively increases from 1 (the input spectrogram) to 32, 64, and 128 across the three stages. To accommodate variable temporal lengths, an Adaptive Average Pooling layer aggregates the final feature maps to a fixed resolution of 4×4. The classification head flattens these features into a 2048-unit vector, which is passed through a fully connected layer of 256 units, a ReLU activation, and a Dropout layer with a 0.5 probability to mitigate overfitting, before the final linear projection to the target classes. This architecture is implemented using the PyTorch framework[38].

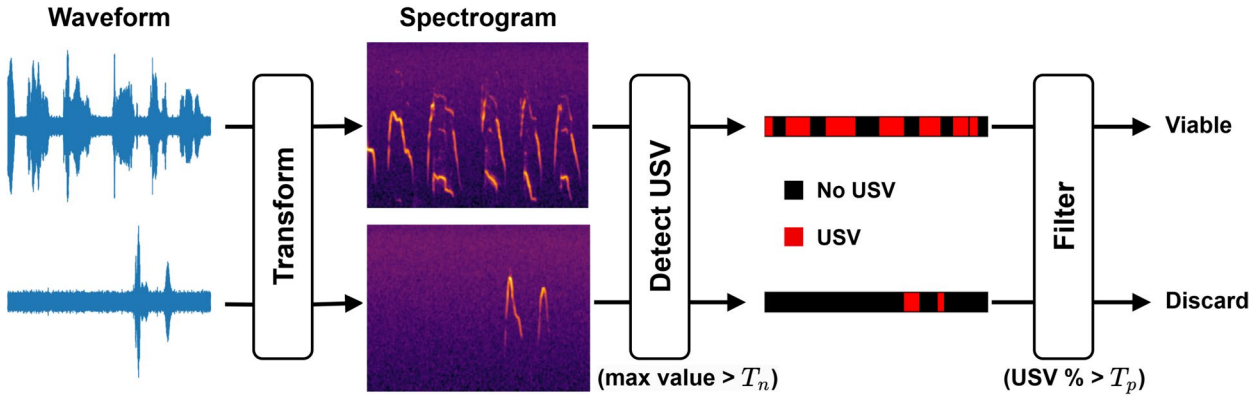

Fig. 3. Pre-processing filtering diagram. Each audio segment is first transformed into the frequency domain to detect USV signals, and the segments with insufficient percentage of USV signals are then discarded.

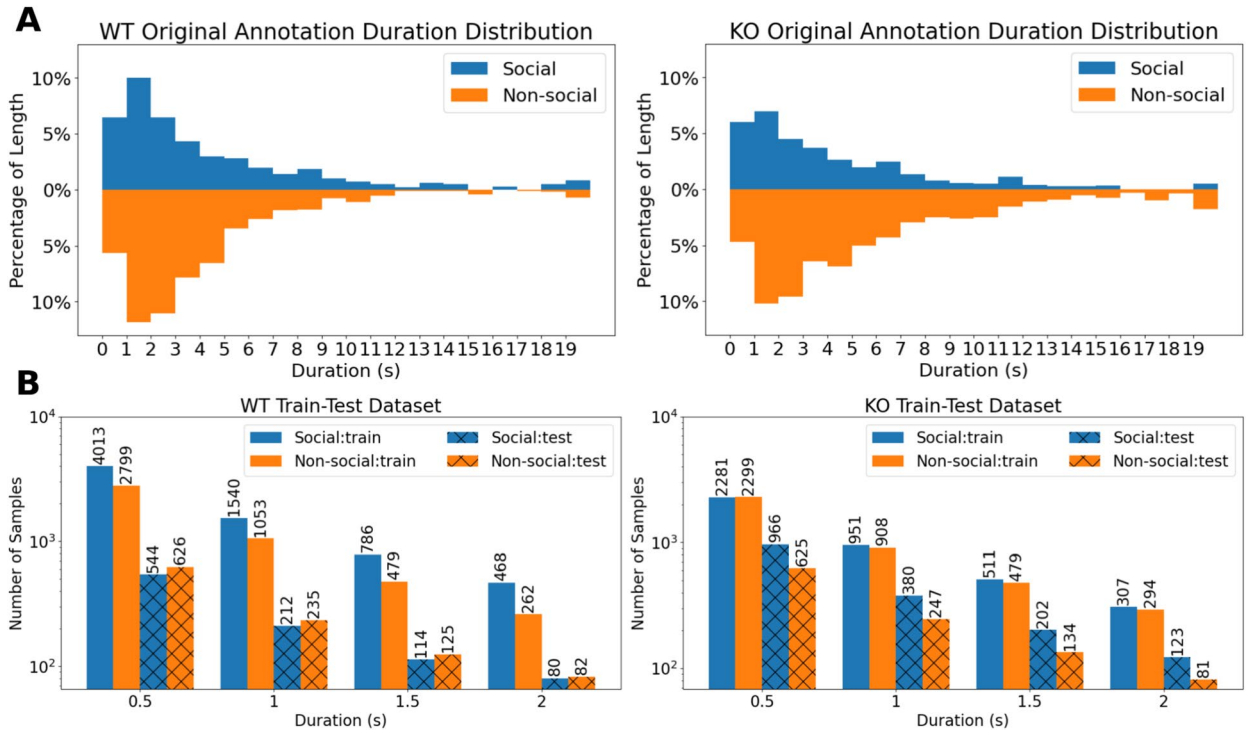

Fig. 4. Distribution of the duration of annotated social and non-social behaviors for WT and KO mice, based on expert human annotations. Such distribution was used to guide the selection of the segment duration adopted for model training. (**B**), Number of training and testing segments in social and non-social behaviors for WT and KO mice after pre-processing. Segment counts directly reflect the subject-level split performed at the recording stage prior to segmentation and pre-processing (≈85% training/validation, ≈15% test).

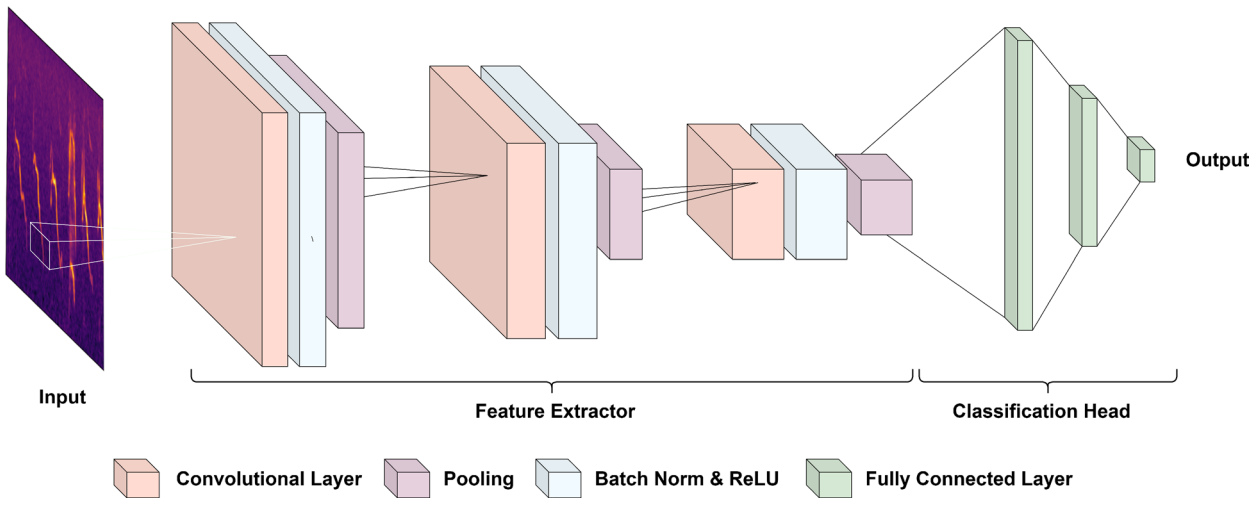

Fig. 5. Model architecture of the proposed DNN classifier. It consists of a feature extraction module with three convolutional layers followed by a classification head with fully connected layers.

Experimental results

Implementation details

To ensure the reproducibility, the model we used was implemented using PyTorch 2.5.0 and TorchAudio of the same version, with a global random seed set to 42. Audio signals were processed at a sample rate of 250 kHz. Mel-spectrograms were extracted using a Fourier transform with 4096 frequency bins, a hop size of 1024 samples between successive frames, and 128 Mel filters, focusing on the 30–125 kHz frequency range. To train the classifier, we employ the Adam optimizer[39], an adaptive optimization algorithm widely used for its stable and efficient convergence in deep learning models. We used a learning rate of 0.0005 and beta parameters set

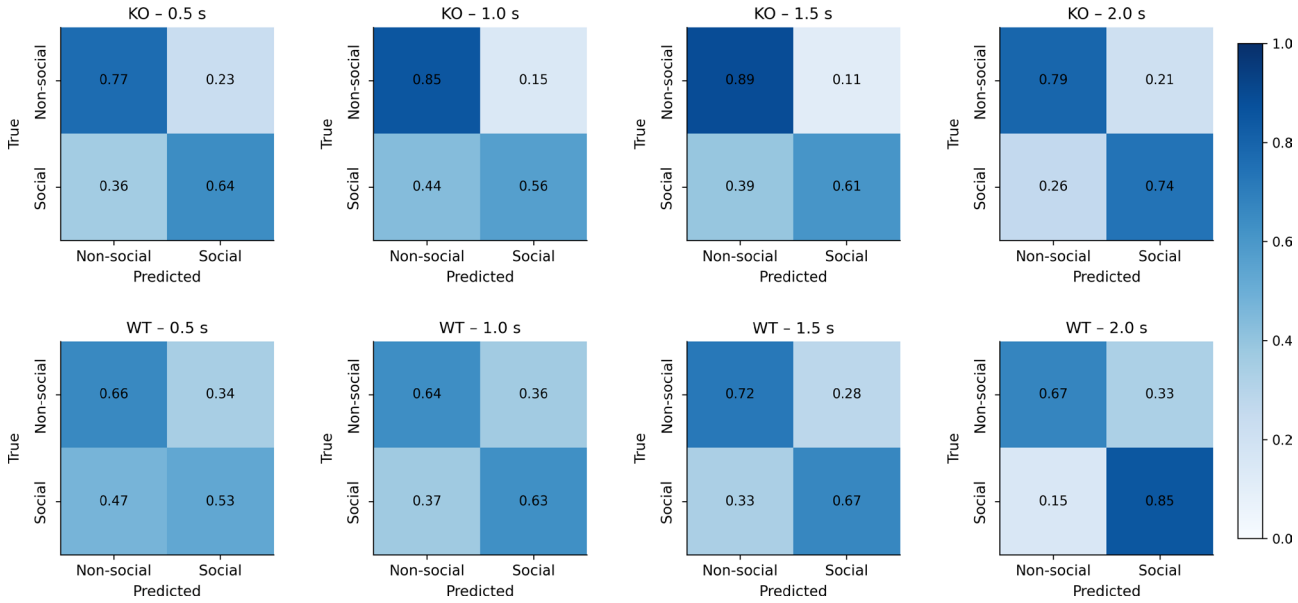

Fig. 6. Confusion matrices of the classification results, with the x-axis representing predicted classes and the y-axis representing ground truth classes. Each entry indicates the percentage of predictions within each ground truth class, with diagonal values showing correct classifications.

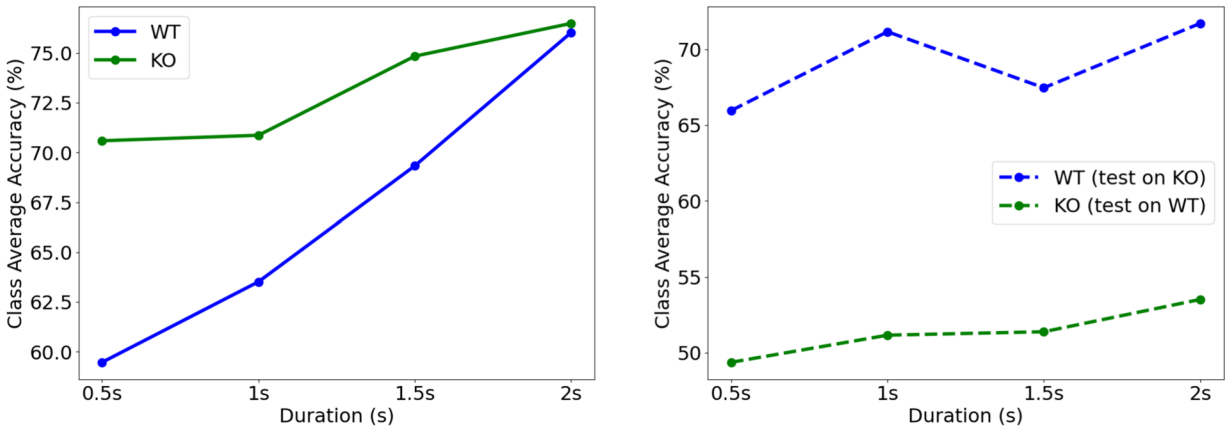

Fig. 7. Class average accuracy across different segment durations. Left panel: performance when each genotype-specific model (WT or KO) is trained and tested on the same genotype. Right panel: cross-genotype performance, where models trained on one genotype are tested on the opposite genotype.

to (0.9, 0.999). We utilized a `ReduceLROnPlateau` scheduler to reduce the learning rate by a factor of 0.5 if the validation loss did not improve for 10 consecutive epochs. Training was conducted with a batch size of 64 for 300 epochs using Cross-Entropy Loss. To address class imbalance in the training set, we employed a `WeightedRandomSampler`, which assigns sampling weights inversely proportional to class frequency. This strategy ensures that minority-class samples are drawn more frequently during mini-batch construction, promoting balanced gradient updates without altering the original data distribution. Data augmentation was applied during training with a probability of $p = 0.5$ for each technique, including time stretching (0.9–1.1×), pitch shifting (± 2 semitones), and random temporal shifting. All code and hardware-specific configurations are open-sourced for further reference.

Performance

Figure 6 presents the confusion matrices for classification results in both KO and WT mice across various segment durations. The class-average accuracy, computed as the mean of the diagonal values of the normalized confusion matrices (i.e., balanced accuracy), is reported in the left panel of Fig. 7. In addition, Table 1 summarizes class-wise recall, precision, and F1-scores for both social and non-social behaviors, together with overall accuracy and balanced accuracy. Strong correlations between USVs and behaviors were observed, with behavior classes (social vs. non-social) accurately predicted using audio alone, achieving average accuracies

Genotype	Duration (s)	Social			Non-social			Accuracy	Bal. Acc
		Recall	Precision	F1	Recall	Precision	F1		
WT	2	0.850	0.716	0.777	0.671	0.821	0.738	0.759	0.760
	1.5	0.667	0.685	0.676	0.720	0.703	0.711	0.695	0.693
	1	0.632	0.612	0.622	0.638	0.658	0.648	0.635	0.635
	0.5	0.531	0.575	0.552	0.658	0.618	0.637	0.599	0.595
KO	2	0.740	0.843	0.788	0.790	0.667	0.723	0.760	0.765
	1.5	0.609	0.891	0.724	0.888	0.601	0.717	0.720	0.748
	1	0.563	0.856	0.679	0.854	0.560	0.676	0.678	0.709
	0.5	0.644	0.811	0.718	0.768	0.583	0.663	0.693	0.706

Table 1. Classification performance for WT and KO mice across segment durations. Metrics are reported symmetrically for both behavioral classes.

reaching up to 76% for 2-second segments. As segment duration increases from 0.5 second to 2 seconds, the classification accuracy improves, suggesting that longer segments provides more comprehensive vocalization patterns, making behavioral distinctions clearer. Notably, the overall class average accuracy was consistently higher in KO mice than in WT mice, particularly for shorter segment lengths. However, a closer inspection of the confusion matrices in Fig. 6 reveals class-specific differences across genotypes and segment durations. In KO mice, non-social behaviors are classified with higher accuracy than social behaviors, whereas in WT mice the performance is more balanced and improves for social behaviors as segment duration increases.

Furthermore, to examine differences in USVs between KO and WT mice, we conducted additional evaluations by switching the testing USV type (Fig. 7, right). The classifier trained on WT USVs performs well even when tested on KO USVs (blue line). However, training on KO USVs and testing on WT USVs results in poor accuracies akin to random guesses, losing the correlation between USV and behavior (green line). These findings indicate an asymmetric cross-genotype generalization pattern: the classifier trained on WT USVs generalizes to KO data, whereas the model trained on KO USVs fails to generalize to WT. This asymmetry suggests that the spectrotemporal variability captured in WT segments encompasses patterns also present in KO mice, while the converse does not hold.

Discussion and conclusions

Our work introduces the first deep learning approach for predicting mouse behavior based solely on their USV communication. By leveraging a DNN classifier, we demonstrated the feasibility of inferring behavior types (social or non-social) from USVs alone. Our findings provide a foundation for future research into the interplay between vocal communication and behavior in mice, with potential applications extending to other animal species.

The study was conducted by analyzing the two genotypes–wild-type (WT) and knockout (KO) mice–both separately and in a cross-genotype setting. Two distinct classifiers were developed, one trained on WT USVs and the other on KO USVs, and each model was subsequently evaluated on both genotypes. When applied to the genotype on which it was trained, each classifier demonstrated a clear correlation between USVs and behavior. Notably, the model trained on WT mice maintained meaningful predictive performance when tested on KO mice. In contrast, the classifier trained on KO mice failed to generalize to WT data, with performance dropping toward chance level. This asymmetric cross-genotype generalization pattern provides further insight into the structure of the vocal representations learned by the network. The fact that WT-trained models generalize to KO mice suggests that the spectrotemporal variability present in WT segments encompasses patterns also observed in KO mice. Conversely, the lack of generalization in the opposite direction indicates that the representations learned from KO data do not fully capture the broader variability exhibited by WT vocalizations. We remark that, in the present study, USV analyses were restricted to male experimental mice, as the resident–intruder male–female paradigm represents a well-established and extensively characterized framework for adult vocal emission; the investigation of female-derived USVs, particularly within female–female interaction settings, will be addressed in future studies.

Importantly, these findings stem from the global acoustic information made available to the model. Rather than relying on predefined descriptors such as call duration, peak frequency, or modulation patterns, the network processes full spectrogram segments and learns directly which aspects of the signal are behaviorally informative. This approach enables the network to identify behaviorally relevant acoustic patterns directly from raw time–frequency inputs, without relying on predefined vocal categories or manually engineered features as in previous studies[9]. Accordingly, our results demonstrate that meaningful behavioral information is embedded in the overall spectrotemporal structure of USVs and can be extracted automatically by an end-to-end model.

This study opens several avenues for future work. One promising direction involves analyzing individual social and non-social behaviors (e.g., anogenital sniffing, nose-to-nose sniffing, physical contact, exploration) separately. While this was not addressed in the current study to maintain analytical simplicity, a more fine-grained behavioral approach could provide deeper insights into the relationship between specific ultrasonic vocalizations and distinct social actions. Such an analysis would require a substantially larger dataset to ensure statistical reliability, particularly for low-frequency behaviors. Although each mouse produces hundreds of USVs

during a 5-minute session, a greater number of subjects would allow for more detailed and robust analyses, enabling stronger generalization.

A further complementary direction concerns the explicit modeling of USV call types. While the present framework focuses on behavior prediction directly from spectrogram segments, vocalization categories themselves carry biologically meaningful information. Our group has previously addressed USV classification using machine and deep learning techniques[16], demonstrating the feasibility of automated call-type identification. Future work could integrate call-type information into behavior prediction. For example, a dedicated segmentation and call-classification module could provide structured vocal features as input to the behavioral model, or a unified multi-stage or multi-task architecture could jointly learn call-type and behavior-level representations.

Ultimately, these future efforts could enhance our understanding of the interplay between vocal communication and behavior, with potential implications for investigating communication in other animal models.

Data availability

The dataset used in the paper, including audio and video clips along with behavioral annotations, is available at https://osf.io/rvkb9.

Code availability

The source code as well as the trained neural network weights are available at https://github.com/JoeK6279/MiceUSVBehavior.

Received: 23 June 2025; Accepted: 19 June 2026
Published online: 15 July 2026

References

1. Holy, T. E. & Guo, Z. Ultrasonic songs of male mice. *PLoS Biol.* **3**, e386 (2005).
2. Yao, K., Bergamasco, M., Scattoni, M. L. & Vogel, A. P. A review of ultrasonic vocalizations in mice and how they relate to human speech. *J. Acoust. Soc. Am.* **154**, 650–660 (2023).
3. Premoli, M., Memo, M. & Bonini, S. A. Ultrasonic vocalizations in mice: relevance for ethologic and neurodevelopmental disorders studies. *Neural Regen. Res.* **16**, 1158–1167 (2021).
4. D'Amato, F. R., Scalera, E., Sarli, C. & Moles, A. Pups call, mothers rush: does maternal responsiveness affect the amount of ultrasonic vocalizations in mouse pups?. *Behav. Genet.* **35**, 103–112 (2005).
5. Panksepp, J. B. et al. Affiliative behavior, ultrasonic communication and social reward are influenced by genetic variation in adolescent mice. *PLoS ONE* **2**, e351 (2007).
6. Peleh, T., Eltokhi, A. & Pitzer, C. Longitudinal analysis of ultrasonic vocalizations in mice from infancy to adolescence: Insights into the vocal repertoire of three wild-type strains in two different social contexts. *PLoS ONE* **14**, e0220238 (2019).
7. Egnor, S. R. & Seagraves, K. M. The contribution of ultrasonic vocalizations to mouse courtship. *Curr. Opin. Neurobiol.* **38**, 1–5 (2016).
8. Asaba, A., Hattori, T., Mogi, K. & Kikusui, T. Sexual attractiveness of male chemicals and vocalizations in mice. *Front. Neurosci.* **8**, 231 (2014).
9. Sangiamo, D. T., Warren, M. R. & Neunuebel, J. P. Ultrasonic signals associated with different types of social behavior of mice. *Nat. Neurosci.* **23**, 411–422 (2020).
10. de Chaumont, F., Lemière, N., Coqueran, S., Bourgeron, T. & Ey, E. Lmt usv toolbox, a novel methodological approach to place mouse ultrasonic vocalizations in their behavioral contexts-a study in female and male c57bl/6j mice and in shank3 mutant females. *Front. Behav. Neurosci.* **15**, 735920 (2021).
11. Pilipenko, T. et al. Exploring ultrasonic communication in mice treated with oil: Audio data processing and correlation study with different behaviours. *Eur. J. Neurosci.* **60**, 4244–4253. https://doi.org/10.1111/ejn.16433 (2024) https://onlinelibrary.wiley.com/doi/pdf/10.1111/ejn.16433.
12. Binder, M. S., Cauley, E. B., Cofsky, N. I. & Lemler, M. O. Neonatal vocalization rate predicts future prosocial behavior in c57 bl/6j mice. *Behav. Brain Res.* **486**, 115560 (2025).
13. Sterling, M. L., Teunisse, R. & Englitz, B. Rodent ultrasonic vocal interaction resolved with millimeter precision using hybrid beamforming. *eLife* **12**, e86126. https://doi.org/10.7554/eLife.86126 (2023).
14. Oliveira-Stahl, G. et al. High-precision spatial analysis of mouse courtship vocalization behavior reveals sex and strain differences. *Sci. Rep.* **13**, 5219 (2023).
15. Coffey, K. R., Marx, R. E. & Neumaier, J. F. Deepsqueak: A deep learning-based system for detection and analysis of ultrasonic vocalizations. *Neuropsychopharmacology* **44**, 859–868 (2019).
16. Premoli, M. et al. Automatic classification of mice vocalizations using machine learning techniques and convolutional neural networks. *PLoS ONE* **16**, e0244636 (2021).
17. Fonseca, A. H., Santana, G. M., Bosque Ortiz, G. M., Bampi, S. & Dietrich, M. O. Analysis of ultrasonic vocalizations from mice using computer vision and machine learning. *Elife* **10**, e59161 (2021).
18. Abbasi, R. et al. Capturing the songs of mice with an improved detection and classification method for ultrasonic vocalizations (bootsnap). *PLoS Comput. Biol.* **18**, e1010049 (2022).
19. Goussha, Y. Hybridmouse: A hybrid convolutional-recurrent neural network-based model for identification of mouse ultrasonic vocalizations. *Front. Behav. Neurosci.* **15**, (2022).
20. Stoumpou, V. et al. Analysis of mouse vocal communication (amvoc): a deep, unsupervised method for rapid detection, analysis and classification of ultrasonic vocalisations. *Bioacoustics* **32**, 199–229 (2023).
21. Baggi, D. et al. Extended performance analysis of deep-learning algorithms for mice vocalization segmentation. *Sci. Rep.* **13**, 11238 (2023).
22. Grilli, M. & Memo, M. Nuclear factor-kb/rel proteins: a point of convergence of signalling pathways relevant in neuronal function and dysfunction. *Biochem. Pharmacol.* **57**, 1–7 (1999).
23. Kucharczak, J., Simmons, M. J., Fan, Y. & Gelinas, C. To be, or not to be: Nf-kb is the answer - role of rel/nf-kb in the regulation of apoptosis. *Oncogene* **22**, 8961–8982. https://doi.org/10.1038/sj.onc.1207230 (2003).
24. Mattson, M. P. & Meffert, M. K. Roles for NF-kB in nerve cell survival, plasticity, and disease. *Cell Death Differ.* **13**, 852–860 (2006).
25. Gutierrez, H. & Davies, A. M. Regulation of neural process growth, elaboration and structural plasticity by NF-kB. *Trends Neurosci.* **34**, 316–325 (2011).

26. Bonini, S. A. et al. Nuclear factor kb-dependent neurite remodeling is mediated by notch pathway. *J. Neurosci.* **31**, 11697–11705 (2011).
27. Zhang, Y. & Hu, W. Nfkb signaling regulates embryonic and adult neurogenesis. *Front. Biol.* **7**, 277–291 (2012).
28. Bonini, S. A. et al. Cortical structure alterations and social behavior impairment in p50-deficient mice. *Cereb. Cortex* **26**, 2832–2849 (2016).
29. Mastinu, A. et al. Melanocortin 4 receptor stimulation improves social deficits in mice through oxytocin pathway. *Neuropharmacology* **133**, 366–374 (2018).
30. Premoli, M. et al. Specific profile of ultrasonic communication in a mouse model of neurodevelopmental disorders. *Sci. Rep.* **9**, 15912 (2019).
31. Casanova, M. F. et al. Minicolumnar abnormalities in autism. *Acta Neuropathol.* **112**, 287–303 (2006).
32. Gaudissard, J. et al. Behavioral abnormalities in the fmr1-ko2 mouse model of fragile x syndrome: The relevance of early life phases. *Autism Res.* **10**, 1584–1596 (2017).
33. Hershey, S. et al. CNN architectures for large-scale audio classification. In *2017 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 131–135 (IEEE, 2017).
34. Nanni, L., Maguolo, G., Brahnam, S. & Paci, M. An ensemble of convolutional neural networks for audio classification. *Appl. Sci.* **11**, 5796 (2021).
35. Caligioni, C. S. Assessing reproductive status/stages in mice. *Curr. Protocols Neurosci.* **48**, A–4I (2009).
36. Petroni, V. et al. Long-term behavioral effects of prenatal stress in the fmr1-knock-out mouse model for fragile x syndrome. *Front. Cell. Neurosci.* **16**, 917183 (2022).
37. Fyke, W. et al. Communication and social interaction in the cannabinoid-type 1 receptor null mouse: Implications for autism spectrum disorder. *Autism Res.* **14**, 1854–1872 (2021).
38. Paszke, A. Pytorch: An imperative style, high-performance deep learning library. *Adv. Neural. Inf. Process. Syst.* **32**, (2019).
39. Kingma, D. P. & Ba, J. Adam: A method for stochastic optimization. In *3rd International Conference on Learning Representations, ICLR 2015, San Diego, CA, USA, May 7-9, 2015, Conference Track Proceedings* (eds Bengio, Y. & LeCun, Y.) (2015).

Acknowledgements
The present study is reported in accordance with ARRIVE guidelines.

Author contributions
Conceptualization: A.G., M.P., S.A.B.; data curation: M.P.; methodology: A.G., M.P., C-H.K.; software and documentation: C-H.K.; supervision: A.G., P.M., R.L., S.A.B.; writing (original draft): A.G., M.P., C-H.K.; all authors reviewed, edited, and approved the manuscript, as well as participated in the discussion and interpretation of the findings.

Funding
This work was partially funded by Fondazione UNIBS.

Declarations

Competing interests
The authors declare no competing interests.

Additional information
Correspondence and requests for materials should be addressed to A.G.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.