nature neuroscience

Technical Report

https://doi.org/10.1038/s41593-024-01649-9

Simple Behavioral Analysis (SimBA) as a platform for explainable machine learning in behavioral neuroscience

Received: 26 June 2023

Accepted: 12 April 2024

Published online: 22 May 2024

Check for updates

Nastacia L. Goodwin[1,2,3], Jia J. Choong[1,4], Sophia Hwang[1], Kayla Pitts[1], Liana Bloom[1], Aasiya Islam[1], Yizhe Y. Zhang[1,2,3], Eric R. Szelenyi[1,3], Xiaoyu Tong[5], Emily L. Newman[6], Klaus Miczek[7], Hayden R. Wright[8,9], Ryan J. McLaughlin[8,9], Zane C. Norville[10], Neir Eshel[11], Mitra Heshmati[1,2,3,12], Simon R. O. Nilsson[1,13] ✉ & Sam A. Golden[1,2,3,13] ✉

The study of complex behaviors is often challenging when using manual annotation due to the absence of quantifiable behavioral definitions and the subjective nature of behavioral annotation. Integration of supervised machine learning approaches mitigates some of these issues through the inclusion of accessible and explainable model interpretation. To decrease barriers to access, and with an emphasis on accessible model explainability, we developed the open-source Simple Behavioral Analysis (SimBA) platform for behavioral neuroscientists. SimBA introduces several machine learning interpretability tools, including SHapley Additive exPlanation (SHAP) scores, that aid in creating explainable and transparent behavioral classifiers. Here we show how the addition of explainability metrics allows for quantifiable comparisons of aggressive social behavior across research groups and species, reconceptualizing behavior as a sharable reagent and providing an open-source framework. We provide an open-source, graphical user interface (GUI)-driven, well-documented package to facilitate the movement toward improved automation and sharing of behavioral classification tools across laboratories.

Behavioral neuroscience requires detailed behavior[1], but the notoriously painstaking process of hand-annotating live or recorded assays poses a substantial bottleneck preventing comprehensive behavioral analysis. The manual approach can be arduous, non-standardized and susceptible to confounds produced by observer drift, long analysis times and poor inter-rater reliability[2–4]. These caveats inhibit the detailed study of complex social repertoires in larger datasets and notably provide lower temporal resolution than most modern methodologies, such as in vivo electrophysiological, fiber photometry and single-cell calcium endomicroscopy recordings[5–8]. Computational neuroethology[4]—the marriage of traditional neuroscience techniques, ethological observation and machine learning—is heralded as one potential solution toward deeper behavioral analysis in more ethologically relevant settings. Furthermore, these data are collected at sampling frequencies that match modern neural recording and manipulation techniques.

The recent rapid development of open-source pipelines for markerless animal pose estimation, which allow for accurate tracking of experimenter-defined body parts in noisy and variable environments[9–12], provides a framework for automated machine-learning-based behavioral analyses (Table 1). Using patterns in animal pose over sliding temporal windows, supervised algorithms are trained to find predefined behaviors of interest. These automated behavioral assessments often exceed human performance[13], increase throughput and

A full list of affiliations appears at the end of the paper. ✉e-mail: sronilsson@gmail.com; sagolden@uw.edu

consistency[14] and reduce human bias and anthropomorphism within scoring[15]. Therefore, open-source pipelines for pose estimation and behavioral analysis are increasingly focused on improving computational accessibility to non-specialists via graphical user interfaces (GUIs), easier installation processes and extensive documentation and tutorials. However, as more laboratories (and manuscript and grant reviewers) adopt these techniques as the de facto expected standard, it is increasingly important to focus on model explainability and behavioral nuance.

Explainability methods in behavioral neuroscience[16] aim to determine why and how machine learning models are coming to conclusions, allowing researchers to (1) standardize behavioral definitions if desired, (2) precisely describe and report specialized, non-standard, behavioral variations and (3) more objectively quantify differences in unsupervised behavioral clusters and scrutinize their biological relevance[17]. More precisely, computing and sharing explainability metrics is an essential step in reconceptualizing behavioral classifiers as objective and shareable reagents akin to the commonly used Research Reagent Identifiers (RRID) system for wet lab reagents. As researchers, we are already expected to report behavioral features such as sex, time of day of testing, light cycles and other experimental details, yet the operational definitions of behaviors themselves are often relegated to one to two sentences in the Methods section. Incorporation of explainability metrics allows for objective and complete reporting of behavioral classifiers, leading to enhanced reproducibility. This is not an argument for field-wide standardization of behavioral algorithms but, rather, an opportunity to precisely and objectively capture and report metrics of computer-aided behavioral analyses between experiments and research groups.

Here we present Simple Behavioral Analysis (SimBA) and introduce accessible tools for validation and explainability of supervised behavioral classifications. SimBA is an open-source, primarily GUI-based program built in a modular fashion to increase non-specialized user access to automated behavioral analysis via supervised machine learning techniques. Two parallel and integrated branches of SimBA allow (1) the generation of non-machine-learning-based descriptive statistics of movement and region of interest (ROI) analyses and (2) supervised machine-learning-based behavioral classification. SimBA is agnostic to the choice of animal species or the number of experimental subjects and has been used to classify fish[18], wasp[19], moth[20], mouse[21–38], rat[39,40] and bird[41] behavior. Furthermore, this approach promotes the wider dissemination of classifiers and associated explainability metrics among research groups and is compatible with new or historical videos annotated by open-source packages such as BORIS[42] and commercial packages such as Noldus Observer or EthovisionXT[43].

Notably, SimBA introduces several machine learning interpretability tools and seamlessly incorporates the application of SHapley Additive exPlanation[44] (SHAP) scores, which is one possible solution to providing explainability and transparency of behavioral classifiers. SHAP is a widely cited open-source and post hoc explainability method that can be applied and compared across any of the current machine learning platforms regardless of pose estimation scheme. Furthermore, Shapley values are widely accepted, understood and under continual development and scrutiny in the greater computer science and artificial intelligence (AI) fields. It has been proposed that explainable AI is critical for the future of neuroscience[16]; and, for a deeper discussion into the usefulness of explainability approaches in computational neuroethology, see Goodwin et al.[17].

Here we highlight the functionality and importance of this approach in multiple datasets to demonstrate the utility of explainability metrics in capturing and describing subtle behavioral variations across sex and environment in freely moving social behavior. Specifically, we examined five behaviors (attack, anogenital sniffing, pursuit, escape and defensive behavior) across males and females in chronic social defeat stress (CSDS) assays and across male rodents in varied contextual environments across resident–intruder (RI) or CSDS assays. To understand the relationship between such behavioral differences between laboratories, we performed SHAP analysis on aggressive social behaviors from different research groups and compared the specific timescales and feature bins characterizing behaviors. We show here that this approach provides quantitative descriptions of behavior, allowing for the use of behavior as an objective shareable reagent.

We present a platform for the rapid frame-by-frame supervised analysis of animal behavior, in conjunction with explainability tools that rapidly and accessibly capture behavioral subtleties often left out of standard behavioral metrics. We propose that behavioral classifiers, in combination with explainability metrics, can be publicly shared and used in the fashion of RRIDs to increase reportability and reproducibility within behavioral neuroscience.

Results

Accessible machine learning for behavioral neuroscientists

SimBA provides accessible machine learning tools to non-specialized users using standard computational hardware via a simple, single-line installation, a GUI and extensive documentation. After raw video acquisition (Fig. 1), users can pre-process their videos (for example, cropping, trimming and changing resolution and contrast) in SimBA before performing pose estimation in their open-source program of choice, such as DeepLabCut[10,45], SLEAP[46] DeepPoseKit[9] and MARS[47]. After GUI-guided import of animal tracking data (Fig. 1a), SimBA calculates relationships between body parts across static and dynamic time windows ('features') that are used to train supervised random forest machine learning classifiers for behavioral predictions (Fig. 1b). SimBA, by default, computes explainable feature representations of movements, angles, paths, velocities, distances and sizes within individual frames and as rolling time window aggregates. To provide flexibility for advanced users, the SimBA library includes a larger battery of runtime-optimized feature calculators covering frequentist and circular statistics, anomaly scores, temporal and spectral analyses, relationships between pose estimation and user-defined ROIs, bounding box methods and other machine learning distribution comparison techniques[18,30,39]. All of these can be deployed within user-defined time windows in use cases where default feature calculators are insufficient. Finally, SimBA is highly flexible and can function as a programmatic platform allowing fully customized user-composed feature extraction classes. These hundreds of features per individual video frame can then be used to train supervised machine learning classifiers or can be fed to trained classifiers that create frame-by-frame predictions of the probability of a behavior of interest occurring.

Training supervised machine learning algorithms requires human annotations of a subset of video frames as either positive or negative for the behavior of interest, which algorithms learn to differentiate using the associated feature values. We streamlined the training set construction process by building in-line behavioral annotation tools, including raw video annotation; machine-assisted annotation where the user verifies annotations produced by a prior behavioral model; and methods for importing existing behavioral labels (Fig. 1b). This process is supported by video batch post-processing tools that can be used to create targeted video clips containing the behaviors of interest that maximize biological replicates and contain similar amounts of positive and negative frames (Fig. 2c), precluding the need to annotate every frame of individual videos. Owing to its ease of adoption for new users, its interpretability and its robustness to overfitting[48–50], our pipeline uses random forest supervised machine learning algorithms; this further allows for the calculation of classical machine learning performance metrics and supports the creation of multiple visualizations and other hands-on validation tools for individual classifiers (Fig. 1b).

Table 1 | Open-source programs for automated behavioral detection

Year	Name	Citation	Validated species	Behavioral classifiers validated in the original publication	Software website
Supervised					
2009	CADABRA	Dankert et al.[78]	Fly	Lunging, tussling, wing threat/extension, circle, chase, copulation	http://www.vision.caltech.edu/cadabra/
2012	MiceProfiler	de Chaumont et al.[79]	Mouse	Head–head contact, anogenital contact, side-by-side, approach, leave, follow, chase	http://icy.bioimageanalysis.org/plugin/mice-profiler-tracker/
2013	JAABA	Kabra et al.[64]	Fly and mouse	Fly: walk, stop, crabwalk, backup, touch, chase, jump, copulation, wing flick/groom/extension, right, center pivot, tail pivot. Mouse: follow, walk.	https://sourceforge.net/projects/jaaba/files/
2013	Unamed	Giancardo et al.[80]	Mouse	Nose to body, nose to nose, nose to genitals, above, follow, stand together	
2015	Unamed	Hong et al.[81]	Mouse	Attack, close investigation, mount	
2020	SimBA	Goodwin et al.[82]	Mouse and rat	Attack, pursuit, lateral threat, anogenital sniff, mount, upright submissive, allogroom, flee, scramble, box, approach, avoidance, drink, clean, eat, rear, walk away, circle	https://github.com/sgoldenlab/simba
2021	MARS	Segalin et al.[47]	Mouse	Attack, mount, close investigation, face-directed, genital-directed	https://github.com/neuroethology/MARS
2022	DeepEthogram	Bohnslav et al.[83]	Fly and mouse	Nose-to-nose, nose-to-body, body-to-body, chase, anogenital	https://github.com/jbohnslav/deepethogram
2022	DeepCaT-z	Gerós et al.[84]	Rodent	Stand still, rear, walk, groom	https://github.com/AnaGeros/Deep-CaT-z-Software
2022	DeepAction	Harris et al.[85]	Mouse	Walk, rest, rear, mover, hang, groom, eat, drink, attack, approach, sniff, copulation	https://github.com/carlwharris/DeepAction
2023	LabGym	Hu et al.[86]	Fly, larva, mouse and rat	Nest build, curl, uncoil, abdomen bend, wing extension, walk, rear, facial groom, sniff, ear groom, sit	https://github.com/umyelab/LabGym
Semi-supervised					
2022	SIPEC	Marks et al.[87]	Mouse and primate	Social groom, search, object interaction	https://github.com/SIPEC-Animal-Data-Analysis/SIPEC
Self-supervised					
2009	Ctrax	Branson et al.[88]	Fly	Walk, stop, sharp turn, crabwalk, backup, touch, chase	http://ctrax.sourceforge.net/
2021	TREBA	Sun et al.[72]	Fly and mouse	Sniff, mount, attack, lunge, tussle, wing extension	https://github.com/neuroethology/TREBA
2022	Selfee	Jia et al.[89]	Fly and mouse	Chase, wing extension, copulation attempt, copulation, social interest, mount, intromission, ejaculation	https://github.com/EBGU/Selfee
Unsupervised					
2014	MotionMapper	Berman et al.[90]	Fly	Groom: wing, leg, abdomen, head. Wing waggle, run	https://github.com/gordonberman/MotionMapper
2017	DuoMouse	Arakawa et al.[91]	Mouse	Sniff, follow, indifferent	
2019	LiveMouseTracker	de Chaumont et al.[79]	Mouse	Rear, look up, look down, contact	https://livemousetracker.org
2020	AlphaTracker	Chen et al.[92]	Mouse	Unnamed individual and social behavior clusters	https://github.com/ZexinChen/AlphaTracker

Table 1 (continued) | Open-source programs for automated behavioral detection

Year	Name	Citation	Validated species	Behavioral classifiers validated in the original publication	Software website
2021	Behavior Atlas	Huang et al.[93]	Mouse	Clusters identified: walk, run, rear, trot, step, sniff, up stretch, left turn, right turn, rise, fall, dive	https://behavioratlas.tech/
2022	VAME	Luxem et al.[94]	Mouse	Motifs identified: exploration, turn, stationary, walk to rear, walk, rear, unsupported rear, groom, backward	https://github.com/LINCellularNeuroscience/VAME
2022	DBscorer	Nandi et al.[95]	Mouse and rat	Immobility	https://github.com/swanandlab/DBscorer
Heuristics					
2022	BehaviorDEPOT	Gabriel et al.[96]	Mouse	Freeze, jump, rear, escape, locomotion, novel object exploration	https://github.com/DeNardoLab/BehaviorDEPOT

Classifier construction and performance

In addition to accommodating shared classifiers and pooled annotation sets, pre-existing classifiers can be adapted to new experimental cohorts and conditions via GUI-assisted thresholding, with limited additional classifier training[34]. Random forest classifiers output a probability per frame of a behavior of interest occurring. As new cohorts of animals are screened and recording contexts are altered, the classifier certainty may decrease for positive frames, but the probability of non-event frames typically stays extremely low. As such, achieving appropriate performance on new samples is supported by dynamic control of discrimination thresholds using an interactive thresholding tool for positive versus negative behavior events (Figs. 1b and 2c). Precise thresholds can also be calculated via precision-recall curves (Fig. 2a). Most laboratories will eventually perform manipulations that alter behavior setups to the point of classifier failure. This is solved by the expansion and incorporation of new training sets. Using the video batch pre-processing interface, new experimental videos may be annotated, manually or with the assisted scoring platform, and easily added to prior models. Thus, new behavioral frames can be rapidly added to the training sets to improve performance, allowing groups to iteratively create updated behavioral classifiers (Fig. 2c).

Standard machine learning performance metrics for all classifiers are provided (Fig. 2a,b and Supplementary Figs. 3 and 4; but see the Discussion section for comments on data leakage). For most classifiers, performance improvement occurs within the first approximately 20,000 positively annotated frames, equivalent to approximately 11 min using standard 30 frames per second (fps) video acquisition. Classifier performance increases with additional annotations, higher clarity of operational definitions and further iterations for targeted misclassification correction. Rat, Caltech Resident-Intruder Mouse (CRIM) and mouse classifiers achieved F1 performance on behavior-present frames of more than 0.91, 0.73 and 0.77, respectively (Fig. 2a,b and Supplementary Figs. 3 and 4). Within SimBA, classifier training includes multiple checks of performance on novel videos to assess generalizability. Using the built-in visualization and validation tools, new unannotated behavioral videos can be analyzed to assess and subsequently adjust performance as necessary. Notably, hand versus machine scoring results—comparing frame-by-frame classifications on independent videos—indicate high classifier performance both within laboratories (attack Pearson's $R^2 = 0.91$ on 16 independent videos; Supplementary Fig. 6) and across laboratories for multiple classifiers (Pearson's $R^2 = 0.936–0.998$ across eight classifiers[26]; Pearson's $R^2 = 0.77, 0.94, 0.98$ (ref. 31); confusion matrix accuracy = 98.6%, 99.3%, 85.0%[30]).

For each classifier, the maximal F1 score is impacted by pose estimation performance, behavior distinctiveness and training set construction. For example, in the case of pose estimation performance, an approximately 4-mm median error in 'tail end' tracking results in failed classification of tail rattles, whereas a 1–2.5-mm median error in 'body hull' points does not affect classifications of other behaviors at F1 > 0.74 (Extended Data Fig. 1). SimBA has several pose estimation outlier correction options available for users upon data import (Extended Data Fig. 2). For behaviors associated with distinct pose estimation signatures, such as drinking, small training sets are sufficient for high performance (<2 min of positive frames, F1 = 0.965). Conversely, to accurately classify multiple behaviors that share similar pose estimation signatures, such as attack and mounting, larger and more diverse training sets are required (attack: >30 min of positive frames, F1 = 0.921; Fig. 2 and Extended Data Fig. 3). Comparing the 'chase' classifier from the CRIM datasets (~2 min of positive frames, F1 = 0.717) and the 'pursuit' classifier from the University of Washington (~1 min of positive frames, F1 = 0.853) reveals the influence of training set construction on F1 score. Ultimately, users will reach an asymptotic point at which further training does not improve classification performance, which depends largely on tracking performance, the set of used features and the distinctiveness of the behavior of interest (Fig. 2a).

SHAP reveals differences among annotators, species and behaviors

Explanations for how machine learning models reach their decisions can help researchers communicate and compare the results of disparate classifiers and support researchers in making informed decisions for machine model implementation and use[51–54]. One recent influential and accessible approach for generating explainable metrics of tree-based classifiers is SHAP[55]. We chose to use SHAP because it is one of the most widely adopted open-source explainability methods available, in addition to being well documented and under active development and support[44]. Shapley values are just one paradigm among a continuously growing number of approaches for explainable artificial intelligence (XAI); alternate options include LIME[52] and counterfactuals[56], all with different pros and cons. Although it is preferable for new pipelines to be built with inherently explainable algorithms[57], we propose that post hoc explainability metrics are highly compatible with the current state of the field owing to the diversity of machine learning packages being developed.

In essence, the Shapley value presents a game-theory-based method for equitably distributing a game's earnings among its players. To illustrate, consider a scenario where individual feature contributions need to be assessed for a binary behavioral classifier.

https://doi.org/10.1038/s41593-024-01649-9

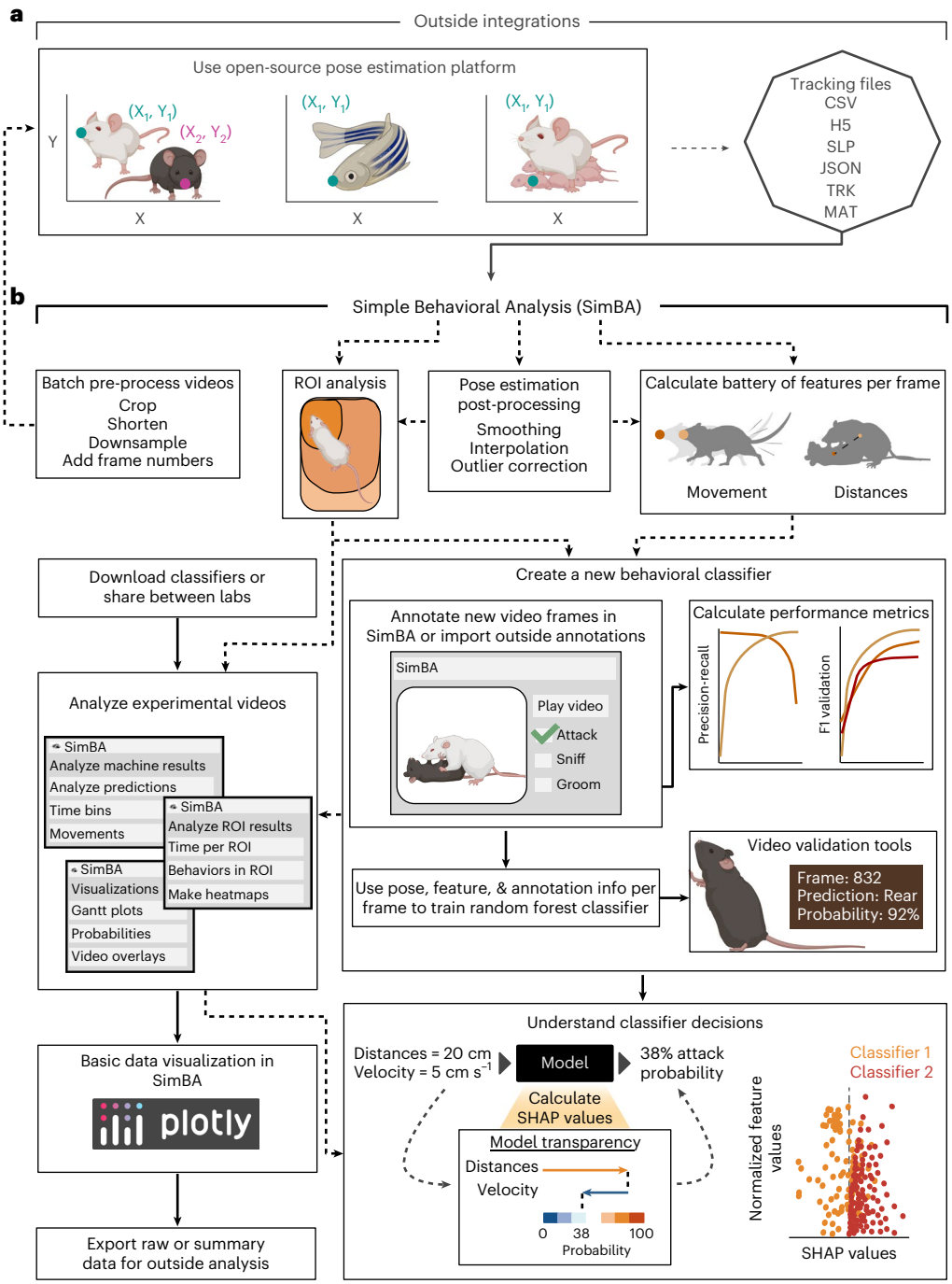
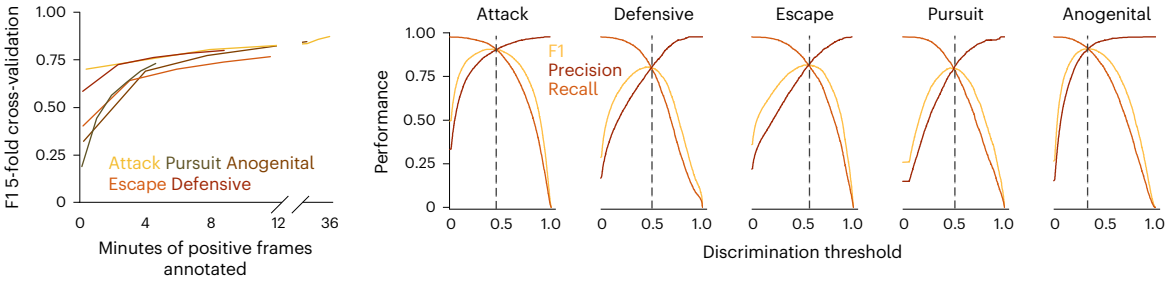


Fig. 1 | SimBA workflow and outside integrations. SimBA is an open-source, GUI-based program built in a modular fashion to address many of the specific analysis needs of behavioral neuroscientists. SimBA contains a suite of video editing options to prepare raw experimental videos for markerless pose tracking, behavior classifications and visualizations. Once users have analyzed their videos for animal pose data via common open-source pipelines (**a**), the data are imported to SimBA for subsequent analysis (**b**). Within SimBA, users have the option to perform pose estimation outlier corrections, interpolation and smoothing methods or to use uncorrected pose data in any SimBA module. To perform supervised behavioral classification, users can download pre-made classifiers from our OSF repository, request classifiers from collaborators or create classifiers by annotating new videos in the scoring interface. Users can also use historical laboratory annotations created in programs such as Noldus ObserverXT, Ethovision or BORIS. A variety of tools are provided for evaluating classifier performance, including calculating standard machine learning metrics, and visualization tools for easy hands-on qualitative validation. After behavioral classification, users can perform batch analyses and extract behavioral measures. To understand the decision processes of classifiers, we encourage users to calculate and report explainability metrics, including SHAP values. We provide extensive documentation, tutorials and step-by-step walkthroughs for all SimBA functionality.

Two key pieces of information are available beforehand. First, there is the average probability (expected value) indicating the likelihood of the behavior occurring in a video frame, such as 10% based on the number of positive annotations from the training data. Second, the model predicts the presence of the behavior in a new frame with 75% probability. By employing Shapley values, the disparity between these values (in this case, 65%) is attributed to the combination of feature values. This difference is then allocated to the features based on their

https://doi.org/10.1038/s41593-024-01649-9

a Machine learning performance metrics



b Classifier information

Classifier	No. of annotated frames	No. of positive frames	Positive annotations (hh:mm:ss)	Under sample	Optimal threshold	F1	Precision	Recall
Attack	553,314	55,300	00:30:43	0	0.46	0.921	0.940	0.904
Defensive	602,606	21,237	00:11:47	17	0.50	0.839	0.847	0.832
Escape	365,822	15,852	00:08:48	10	0.56	0.832	0.790	0.882
Pursuit	470,229	8,268	00:04:35	12	0.51	0.839	0.853	0.826
Anogenital sniffing	910,941	27,613	00:15:20	0	0.36	0.892	0.988	0.892

c Classifier construction

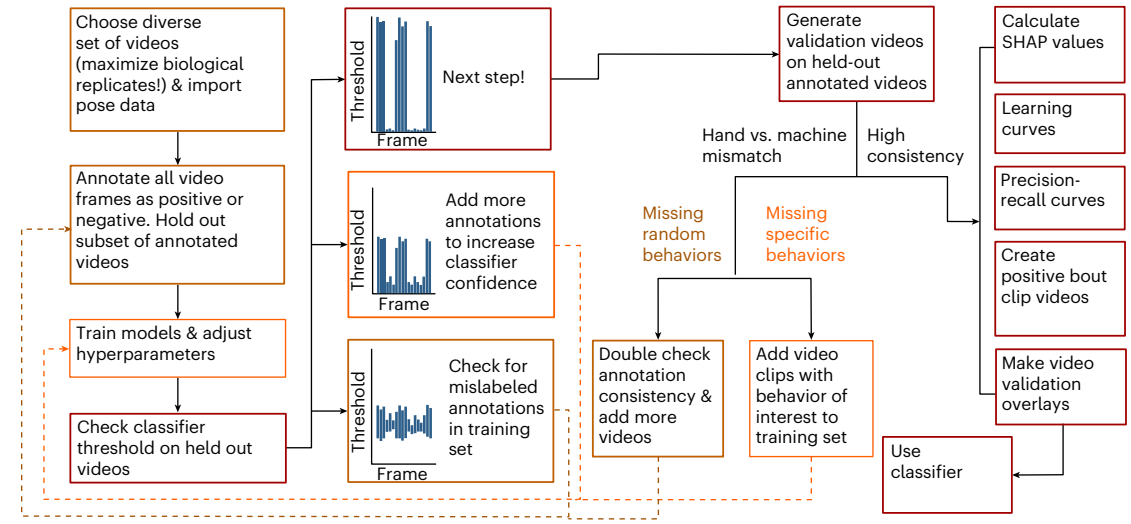


Fig. 2 | Classifier construction workflow and classifier performance metrics. **a**, Machine learning performance metrics for the classifiers used in Figs. 5 and 6 (see Extended Data Fig. 3 and Supplementary Figs. 1–6 for in-depth classifier performance data). Left, F1 five-fold cross-validation learning curves plotted against minutes of positive frames annotated (30 fps). Right, precision-recall curves plotted against discrimination threshold for five classifiers, which can be used in combination with the SimBA interactive thresholding visualization tool to determine the most appropriate detection threshold for classifiers and specific datasets. **b**, Extended information for the training sets for each of the five classifiers. **c**, Workflow for creating high-fidelity and generalizable supervised behavioral classifiers. The dotted lines indicate optional loops for iteratively improving classifier performance. Behavioral operational definitions and classifier SHAP values are shown in Supplementary Figs. 1–4.

contributions within the ensemble. The resulting Shapley value output is a vector of the same length as the number of features, with a total sum of 65 (see the Methods section for a detailed mathematical explanation of SHAP and its application in behavioral neuroscience).

SimBA calculates many hundreds of features (Supplementary Table 1) that, individually, may not be very informative. Although SHAP values can be calculated for each of these features if desired, the additivity axiom allows users to bin features into ethologically relevant feature categories (Supplementary Table 2) that are related to the biology of their model system and/or behaviors of interest. Another key advantage is that classifiers generated for the same behavior, but using different pose estimation schemas (for example, one model with five body parts versus a model with eight body parts), may still be directly compared by such bins, which would not be possible using individual feature importances (Extended Data Fig. 4). SHAP is built into SimBA classifier construction to encourage its use and accessibility, but, more so, we propose that the adoption of any type of explainability paradigm is more important than the specific algorithm selected.

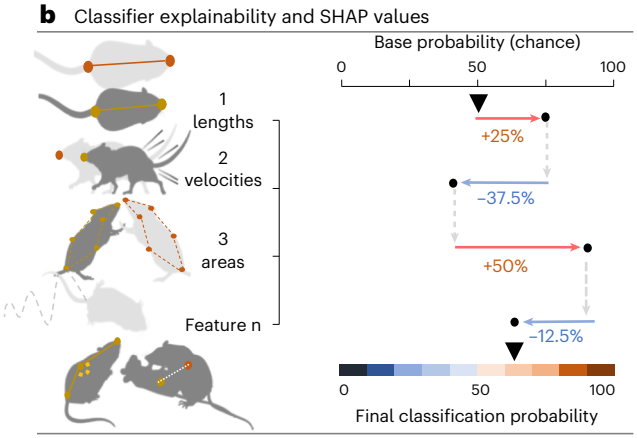
To specifically demonstrate the potential of SHAP to enhance behavioral reportability within preclinical behavioral neuroscience, we collected and compared independently created rodent attack classifiers from expert annotators at institutions across the United States. These classifiers and annotators used data from different recording environments, with different experimental protocols, strains, sex, video formats and pose estimation models (Fig. 3a and Statistical Supplementary Note 1—detailed statistics). The classifiers all showed high accuracy in their respective environments.

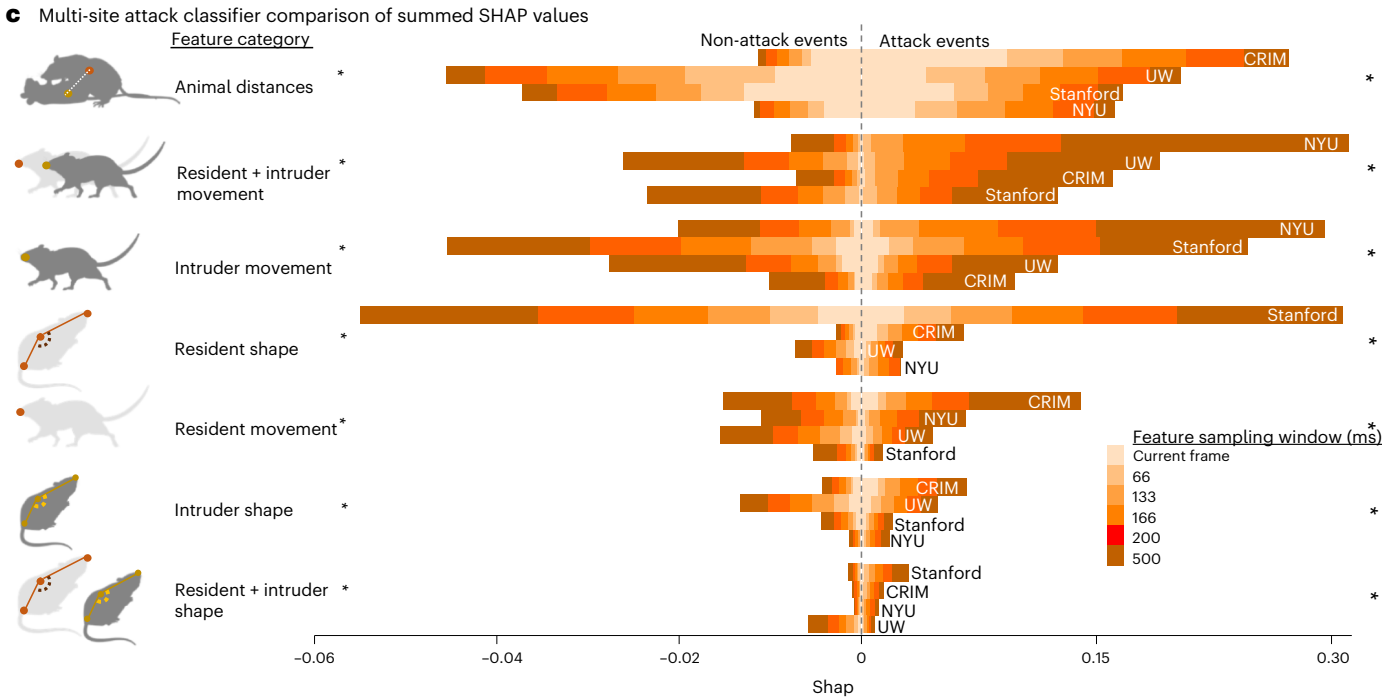
SHAP analysis of attack classifiers revealed the relative importance of feature category bins[17] for discerning both attack and non-attack frames. Across laboratories, animal distances and movement were

https://doi.org/10.1038/s41593-024-01649-9

a Multi-site supervised classification consortium

	Site 1	Site 2	Site 3	Site 4	Site 5	Site 6
Location	UW	Tufts	Stanford	Caltech	NYU	WSU
Group	Golden	Miczek	Eshel / Malenka	CRIM / Perona	Lin	McLaughlin
Species	Mouse	Mouse	Mouse	Mouse	Mouse	Rat
Environment	Home cage	Home cage	Operant	Home cage	Home cage	Home cage
Resident	CD1	CFW	C57	CD1	CD1	Long-Evans
Intruder	C57	C57	BALB/c	C57	C57	Sprague
Sex	♂	♀	♂	♂	♂	♂
Video format	RGB	Grayscale	RGB	Grayscale	RGB	RGB
frame rate	30	30	30	25	30	30
No. of annotations	183,572	326,003	89,564	757,350	464,610	136,830





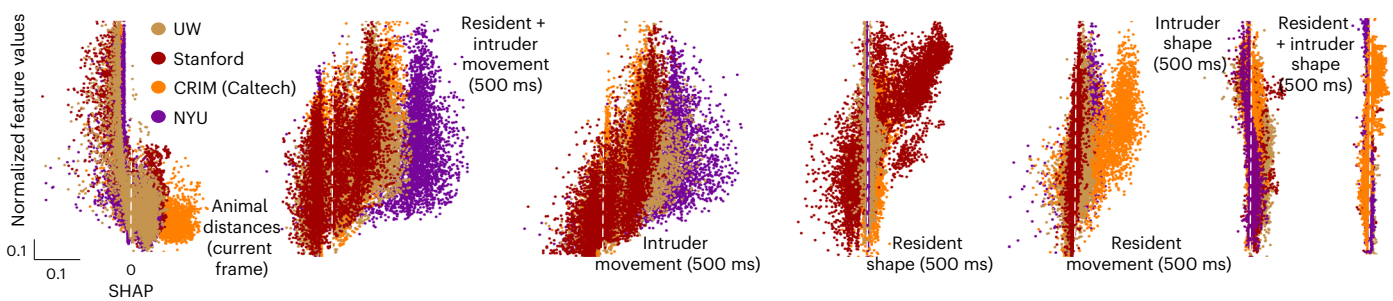


Fig. 3 | SHAP attack classifier consortium data. a, Description of the consortium dataset used for the cross-site attack classifier comparisons. **b**, Schematic description of SHAP values, where the final video frame classification probability is divided among the individual features according to their contribution. **c**, ANOVA comparison of summed feature SHAP values, collapsed into seven behavioral feature categories for four different mouse attack classifiers. We divided each category into six further subcategories that represented features within the categories with different frame sampling frequencies (one frame, 500 ms) and are denoted by shaded colors. Asterisks denote significant main effect of consortium site, $P < 0.0001$. See Supplementary Note 1—detailed statistics for full statistical analysis. **d**, Scatter plots showing the directional relationships between normalized feature values and SHAP scores in four mouse RI attack classifiers and seven feature subcategories. The dots represent 32,000 individual video frames (8,000 from each site's dataset), and color represents the consortium site where the annotated dataset was generated. All tests were two-sided. Bonferroni's test was used for multiple comparisons where applicable. NYU, New York University; RGB, red, green and blue; WSU, Washington State University; UW, University of Washington; n, additional feature bins of interest.

https://doi.org/10.1038/s41593-024-01649-9

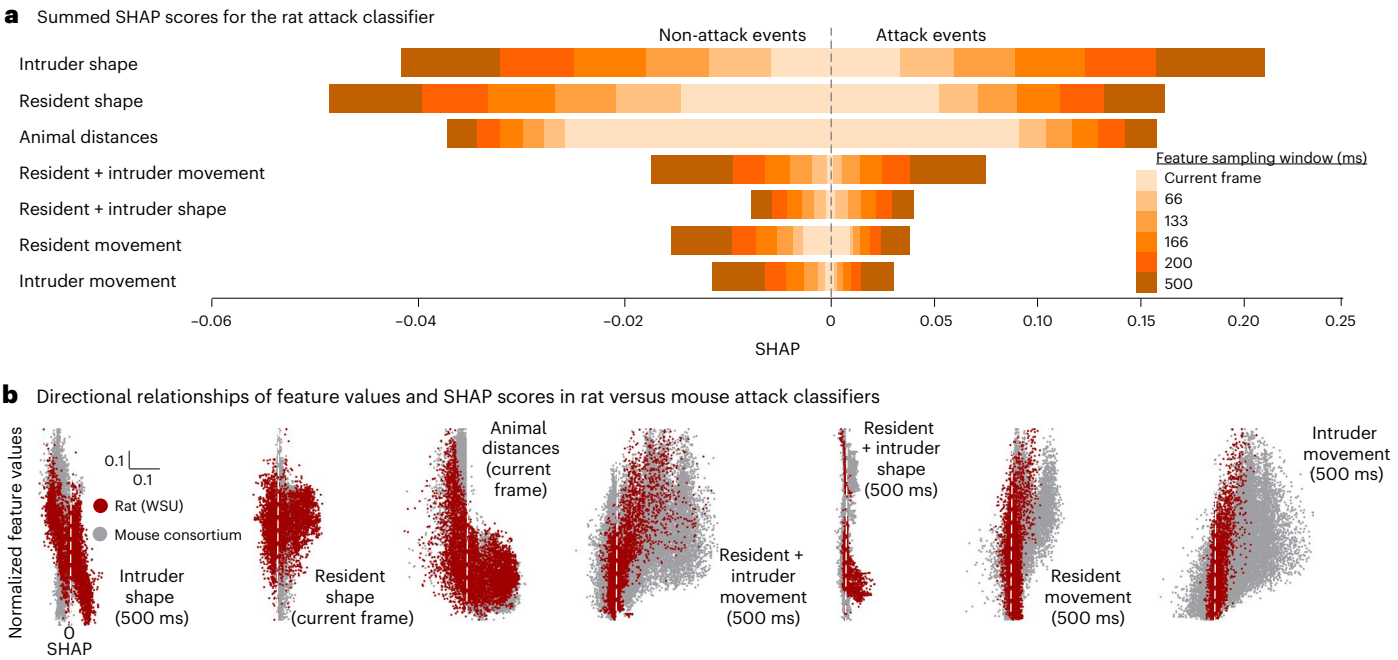


Fig. 4 | SHAP cross-species attack classifier data. Explainable classification probabilities in the rat RI attack classifier using SHAP. **a**, Summed SHAP values, collapsed into seven behavioral feature categories for the rat random forest attack classifier. Colors denote sliding window duration as in Fig. 3. **b**, Scatter plots showing the directional relationships between normalized feature values and SHAP scores in seven feature subcategories of the rat RI attack classifier. The rat attack classifier is shown in red. For comparison, the SHAP values for the mouse attack classifiers (from Fig. 4) are shown in gray. Dots represent individual video frames. See Supplementary Note 1—detailed statistics for full statistical analysis. WSU, Washington State University.

important for identifying attack, whereas animal shapes were least important. In addition to identifying these common patterns, a stark difference in the influence of resident shape was seen between the Stanford University attack classifier and classifiers from other sites (Fig. 3c). To ensure that this was not due to incongruent operational definitions of attack behavior, we manually annotated the Standford University dataset based on our operational definition of attack. Manual annotations were highly consistent between sites (Extended Data Fig. 5; $R^2 = 0.998$; Gantt plot). SHAP analysis showed that University of Washington annotations relied on resident body shape over longer durations than Stanford. This demonstrates that, although the annotators have high inter-rater reliability, SHAP determined that they rely on different features for attack scoring. This observation is likely based on the biological differences between the two site's datasets, where Stanford used C57 resident mice rather than outbred strains like other research locations.

Because SHAP values do not convey directional relationships, we plotted the 32,000 observations within each category (8,000 from each site) with the normalized feature values on the *y* axis and SHAP values on the *x* axis (Fig. 3d and Statistical Supplementary Note 1—detailed statistics). Here, the left-most panel shows that, as the distance between animals decreases, the attack classification probability increases across the sites, with shorter animal distances having a stronger positive impact on attack classification probabilities at CRIM/Caltech than the classifiers annotated at other sites.

Rat versus mouse attack behavior

Next, we analyzed SHAP values for the rat RI classifier and compared the values with the five classifiers generated by the mouse consortium of attack classifiers (Fig. 4a and Extended Data Fig. 6). Rat attack events were recognized primarily by the shape of the resident and the intruder and distances between the resident and the intruder. Rat attacks differed significantly by feature category bin but not by feature sliding window size. All feature bin correlations were significant, with intruder shape being most negatively correlated with attack probability in rats and intruder movement being most positively correlated in mice (Fig. 4b).

Automated behavioral analysis of reactive aggression by sex

Female assays for the study of aggression in mice have historically focused on maternal aggression owing to the stated low propensity for these mice to display aggression during reactive aggression tasks (RI or CSDS tests) typically used for males[58]. Recently, studies have revisited reactive aggression in females and have found that a subset of female Swiss Webster (CFW) mice attack same-sex intruders when either virgin and singly housed[59] or co-housed with a castrated male[60]. Although CFW females do not appear to show aggression reward by standard measures, including aggression conditioned place preference and aggression self-administration[61], this still presents an opportunity to directly compare male and female reactive aggression behavior.

After screening females for RI aggression, we conducted female[60,61] and male[62] CSDS assays (Methods). We calculated the total duration, number of bouts, mean bout duration and interval for attack, pursuit, anogenital sniffing, escape and defensive behaviors (SHAP values; Extended Data Fig. 7), both per day of testing, and averaged across all five testing days. We analyzed attack, pursuit and anogenital sniffing for the residents and defensive and escape behavior for the intruders. Males and females showed significant differences in all five assayed behaviors (Fig. 5d), with females showing higher average total durations and number of bouts in attack, pursuit and escape behaviors ($P < 0.001$ for all), whereas males had higher levels of anogenital sniffing and defensive behaviors (duration: $P = 0.0461$, 0.0374; bouts: $P = 0.0298$, 0.0146). No differences were observed

https://doi.org/10.1038/s41593-024-01649-9

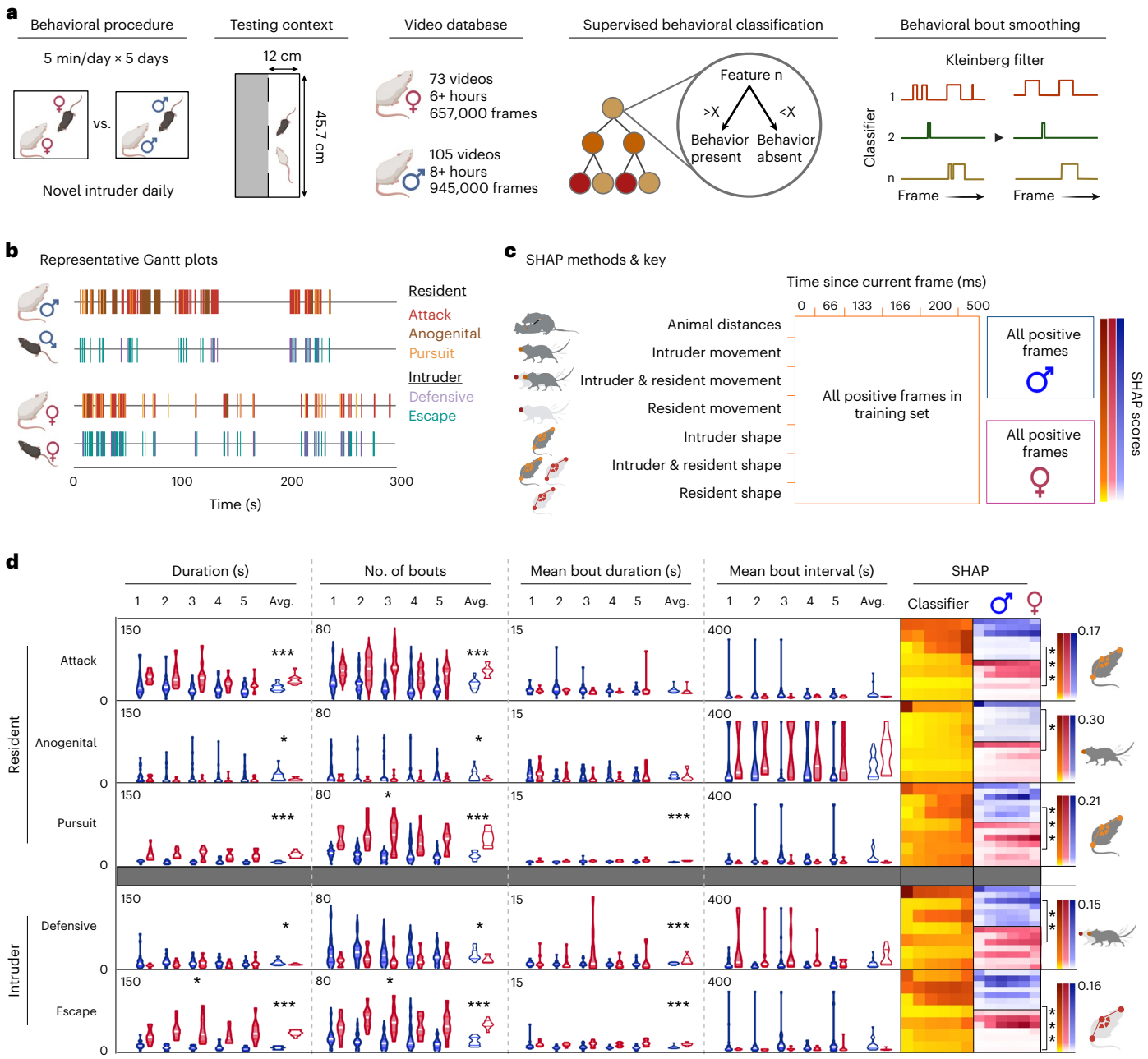


Fig. 5 | Social stress experience influences aggression and coping behaviors differently in males and females. a, Schematic representation of the mouse CSDS behavioral protocol and the analysis pipeline for supervised machine learning behavioral classification. **b**, Representative Gantt charts of classified male (top) and female (bottom) resident and intruder behaviors. **c**, Key for the SHAP analysis and feature bin comparisons. **d**, Supervised behavioral data and SHAP comparisons for five behavioral classifiers. We analyzed attack, pursuit and anogenital sniffing for the residents and defensive and escape behavior for the intruders. Male data are represented in blue and female data in pink. For each classifier, SimBA provided the total duration (s), number of classified bouts, mean bout duration (s) and mean bout interval (s) across individual testing days ($n = 21$ males for all classifiers; $n = 11$ female residents for attack, anogenital sniffing, pursuit and classifiers and $n = 10$ female intruders for defensive and escape classifiers). Males and females showed significant differences in all five assayed behaviors, with females showing higher average total durations and number of bouts in attack, pursuit and escape behaviors ($P < 0.001$ for all), whereas males had higher levels of anogenital sniffing and defensive behaviors (duration: $P = 0.0461, 0.0374$; bouts: $P = 0.0298, 0.0146$). Only three metrics were significantly affected by day: escape duration and bouts (duration: interaction $P = 0.0122$, day $P = 0.3700$, sex $P < 0.001$; bouts: interaction $P = 0.0415$, day $P = 0.3947$, sex $P < 0.001$) and number of pursuit bouts (interaction $P = 0.0404$, day $P = 0.1724$, sex $P < 0.001$). Average SHAP values are reported in Supplementary Fig. 2. The color intensity for all three SHAP datasets per classifier are on the same scale, as indicated by the scales on the right. For each behavior, comparisons with the lowest P value per category are highlighted via a comparison bracket and the feature bin mouse icon. Asterisks denote significance levels: $^*P < 0.05$, $^{**}P < 0.01$ and $^{***}P < 0.001$. See Supplementary Note 1—detailed statistics for full statistical analysis. n, additional classifiers of interest.

for any of the behaviors in mean bout interval or in attack or anogenital mean bout duration. Females showed longer behavioral bouts for pursuit, defensive and escape behaviors ($P < 0.001$ for all). Only three metrics were significantly affected by day: escape duration and bouts (duration: interaction $P = 0.0122$, day $P = 0.3700$, sex $P < 0.001$; bouts: interaction $P = 0.0415$, day $P = 0.3947$, sex $P < 0.001$) and number of pursuit bouts (interaction $P = 0.0404$, day $P = 0.1724$, sex $P < 0.001$).

https://doi.org/10.1038/s41593-024-01649-9

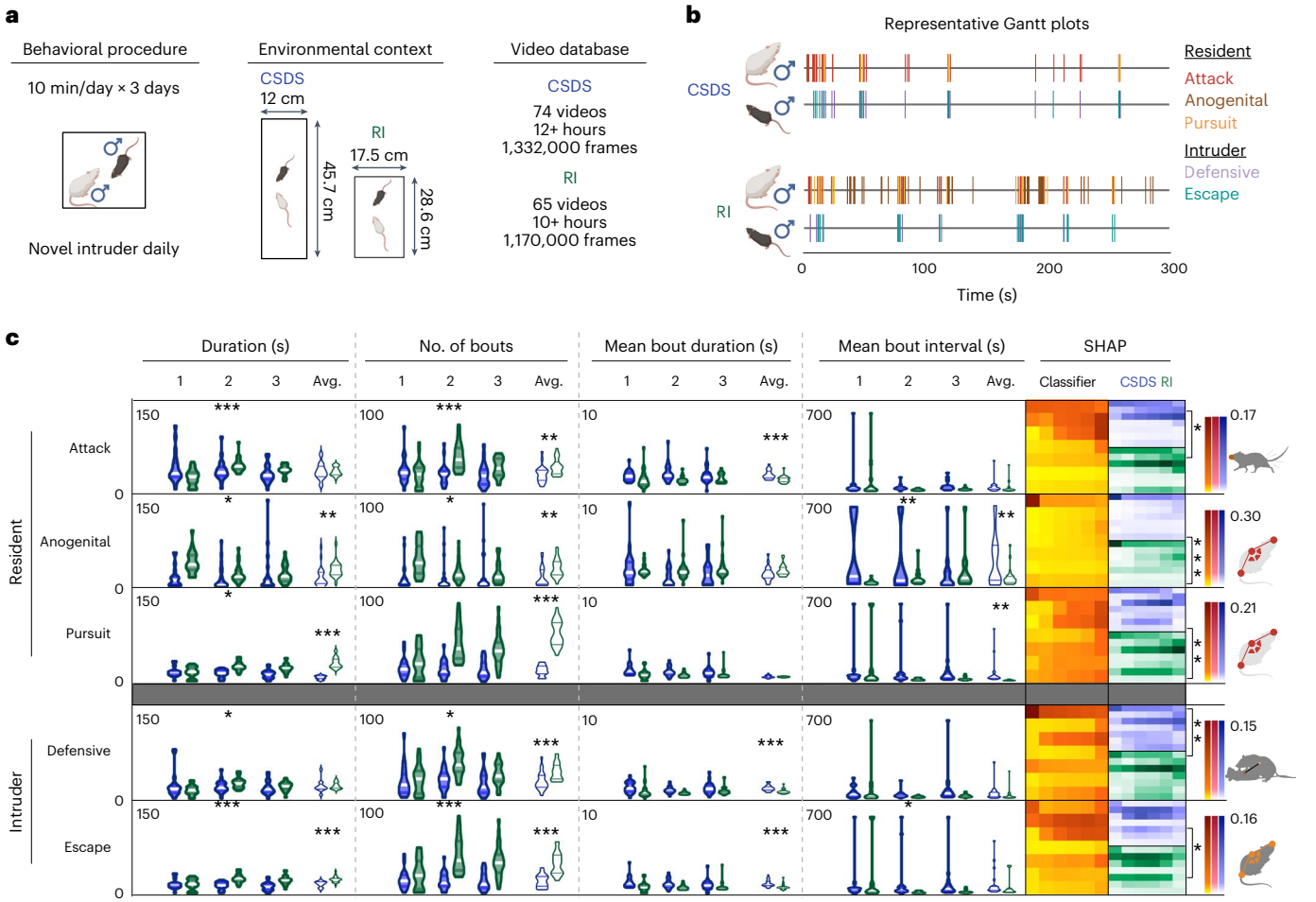


Fig. 6 | Environment and experience influence male aggression and coping behaviors. a, Schematic representation of the mouse CSDS and RI behavioral design. **b**, Representative Gantt charts of classified CSDS (top) and RI (bottom) resident and intruder behaviors. **c**, Supervised behavioral data and SHAP comparisons for five behavioral classifiers. CSDS data are represented in blue, and RI data are shown in green. For each classifier, SimBA provided the total duration (s), number of bouts, mean bout duration (s) and mean bout interval (s) across individual testing days ($n = 21$ CSDS, $n = 24$ RI). RI males showed a marked decrease in anogenital sniffing duration across days (interaction $P = 0.0123$, day $P = 0.0478$, environment $P < 0.0071$), with concomitant increases in attack behavior (interaction $P < 0.001$, day $P = 0.0204$, environment $P = 0.9408$) and pursuit behavior (interaction $P < 0.0258$, day $P = 0.0295$, environment $P < 0.001$). The color intensity for all three SHAP datasets per classifier are on the same scale, as indicated by the scales on the right. For each behavior, comparisons with the lowest P value per category are highlighted via a comparison bracket and the feature bin mouse icon. Asterisks denote significance levels: $^{*}P < 0.05$, $^{**}P < 0.01$ and $^{***}P < 0.001$. See Supplementary Note 1—detailed statistics for full statistical analysis.

To understand how males and females performed behaviors predicted by the same supervised behavioral classes, we calculated SHAP values for 10,000 positive behavioral frames per behavior and sex (Fig. 5d, right). Significant differences were observed in SHAP values across sexes for all behaviors (Supplementary Note 1—detailed statistics), with attack and pursuit behaviors differing most significantly in intruder shape, anogenital sniffing in intruder movement, defensive behavior in combined animal movement and escape behavior in resident shape (Supplementary Note 1—detailed statistics).

Automated analysis of male aggression by environment

RI assays are typically conducted in an animal's home cage to measure the reactive or territorial aggression of the resident animal. Frequently, these assays examine resident males across their first exposures to established subordinate intruders. Alternately, CSDS testing is typically performed in thinner subdivided hamster cages, where the 'resident' animal is an established aggressor[62] and the intruder is a smaller mouse that experiences up to 10 consecutive days of defeat. Directly comparing RI and CSDS datasets allows us to gain a preliminary understanding of the intersection of aggression experience and testing environment for driving social behaviors.

RI and CSDS male behavior differed significantly by day; by total duration across all five behaviors; by number of bouts in attack, anogenital, defensive and escape behaviors; and by mean bout interval between anogenital and escape events (Fig. 6c and Supplementary Note 1—detailed statistics). RI males showed a marked decrease in anogenital sniffing duration across days (interaction $P = 0.0123$, day $P = 0.0478$, environment $P < 0.0071$), with concomitant increases in attack behavior (interaction $P < 0.001$, day $P = 0.0204$, environment $P = 0.9408$) and pursuit behavior (interaction $P < 0.0258$, day $P = 0.0295$, environment $P < 0.001$), indicating a shift from exploratory to aggressive behaviors as males gained aggression experience. SHAP comparisons of classifiers revealed differing behavioral motifs between environments. Attack differed most significantly in intruder movement, whereas anogenital and pursuit behaviors differed in resident shape, defensive behavior by animal distances and escape by intruder shape (Fig. 6c, Supplementary Note 1—detailed statistics and Extended Data Fig. 8).

Discussion

There is a vibrant and growing ecosystem of computational behavioral tools designed specifically for the behavioral neuroscience

community that have directly impacted scientific directions and methods[63] (see Table 1 for a non-comprehensive list of examples). SimBA has been heavily influenced by many of these packages[9,10,46,64] and now influences others[39,65,66]. Since the release of SimBA, numerous independent laboratories have used it to address varied scientific questions across diverse model systems, providing critical feedback on SimBA and its ongoing development. The key feedback has consistently focused on introducing easily accessible explainability metrics that are useful to non-specialized users and allow for generalizable use of classifiers and compatibility of disparate datasets.

In the present study, we used SimBA to analyze data, including male and female CSDS behavior and male CSDS and RI behaviors, and describe the use of explainability metrics. Between sexes, we found that male and female intruders differ significantly in their coping strategies, with females defending themselves less than males. These differences may partially explain differences in resilience to social stress between males and females. Between CSDS and RI assays, which were performed in differently sized arenas, we found differences in resident aggressive and intruder coping behaviors between males. This demonstrates clear cross-protocol behavioral differences and emphasizes a need for standardization and quantification of experimental variation within social behavior research in general and aggression research specifically, and we provide a tool to do this.

We demonstrate the utility of SHAP values for uncovering differences within supervised machine learning scenarios. We also quantify significant differences in features and time bins defining female and male aggression as well as aggressive behaviors across distinct testing environments. For example, SHAP values revealed and quantified potential cross-laboratory classifier confounds and aided comparisons between male and female attack events.

Although strong arguments remain for using inherently explainable models in machine learning[57], many maturing behavioral neuroscience tools still rely on black box methods. The post hoc nature of SHAP allows independent laboratories an algorithmic freedom while maintaining options for cross-model explainability comparisons through game theory and the additivity axiom. A substantial drawback of SHAP, however, is the computational and runtime cost associated with analyzing larger datasets. SimBA depends on the TreeSHAP[67] algorithm to calculate SHAP values. A challenge when calculating SHAP values through TreeSHAP is the computational complexity and runtimes that scale polynomially with the number of estimators and features and can interfere with the analyses of larger datasets[68]. To negate this, SimBA includes multi-processing options that allow users to linearly reduce their runtimes with the count of their available CPU cores. SHAP also represents a post hoc approach when inherently explainable models (for example, using a few carefully selected relevant features joined with low algorithm complexity) may suffice for low-complexity behaviors[17]. Furthermore, as a correlative approach, users should avoid interpreting causal relationships between SHAP-inferred feature importances and classification probabilities. Ongoing efforts in SimBA involve parallelization and just-in-time compilation, making the full catalog of methods available on standard behavioral neuroscience hardware at the maximum speeds allowed by the machine. Notably, the SimBA platform has additional in-built access to other prevalent model interpretability methods (for example, partial dependencies, permutation importances and gini/entropy-based measures) when needed.

Manual behavioral analyses typically depend on non-descript qualitative operational definitions, posing challenges for standardization across individuals and laboratories, and potentially contributing to issues with reproducibility downstream. In contrast, as demonstrated here, the incorporation of machine-learning-based behavioral analysis coupled with explainability metrics produce comprehensive quantitative operational definitions of behavior. This allows us to re-conceptualize behavioral analysis though precise and verbalizable statistical component rules that are applied when scoring behaviors of interest. These quantified definitions can be shared as resources, akin to RRID-like reagents, enhancing transparency and reportability and facilitating cross-study comparisons and reproducibility.

As a supervised learning tool, a weakness of our approach is the cost of collecting the human behavioral annotations required for fitting reliable downstream models. Other prominent platforms may circumvent such costs through active learning, task programming or semi-supervised statistical techniques[69–72]. Despite our publicly available classifiers and annotations, users will typically have to add a few short, varied and representative annotations to their classifier training sets owing to the inherent variation in individual laboratory experimental setups. However, the behavioral neuroscience community has an exceptionally rich historical bounty of carefully documented animal behaviors in video recordings. To lessen user burden, SimBA includes several tools to accommodate use of such historical annotations for supervised machine learning purposes. Additional parallel initiatives, including MABe[73,74], the OpenBehavior project[75] and The Jackson Laboratory Mouse Phenome Database[76], may also address these limitations through the distribution of larger repositories encompassing diverse video recordings with associated human annotations.

A further challenge for the field is appreciating how to accurately judge the performance of machine learning models, which can be biased toward subsets of animals and recording environment and that depend on timeseries data that are vulnerable to leakage[77]. Maintaining true independence across training and hold-out sets is a standing challenge within machine learning that can be particularly difficult in behavioral neuroscience use cases where insufficient amounts of fully independent annotations are available. Fundamentally, although low performance metrics are indicative of an inadequate classifier, high metrics may not be sufficiently indicative of reliable and generalizable out-of-sample performance. We recommend that users include smaller held-out and hand-scored representative validation sets for validation purposes.

In conclusion, SimBA provides a user-centric, modular and accessible platform for machine learning analysis of behavior with the primary aim of promoting an understanding of the behavioral nuances that we value as neuroethologists. Here we demonstrate the utility of SimBA as a behavioral analysis tool by consolidating cross-site behavioral datasets with machine learning methods to quantify and describe complex behaviors across experimental contexts.

Online content

Any methods, additional references, Nature Portfolio reporting summaries, source data, extended data, supplementary information, acknowledgements, peer review information; details of author contributions and competing interests; and statements of data and code availability are available at https://doi.org/10.1038/s41593-024-01649-9.

References

1. Krakauer, J. W., Ghazanfar, A. A., Gomez-Marin, A., MacIver, M. A. & Poeppel, D. Neuroscience needs behavior: correcting a reductionist bias. *Neuron* **93**, 480–490 (2017).
2. Anderson, D. J. & Perona, P. Toward a science of computational ethology. *Neuron* **84**, 18–31 (2014).
3. Egnor, S. E. R. & Branson, K. Computational analysis of behavior. *Annu. Rev. Neurosci.* **39**, 217–236 (2016).
4. Datta, S. R., Anderson, D. J., Branson, K., Perona, P. & Leifer, A. Computational neuroethology: a call to action. *Neuron* **104**, 11–24 (2019).
5. Falkner, A. L., Grosenick, L., Davidson, T. J., Deisseroth, K. & Lin, D. Hypothalamic control of male aggression-seeking behavior. *Nat. Neurosci.* **19**, 596–604 (2016).
6. Ferenczi, E. A. et al. Prefrontal cortical regulation of brainwide circuit dynamics and reward-related behavior. *Science* **351**, aac9698 (2016).

7. Kim, Y. et al. Mapping social behavior-induced brain activation at cellular resolution in the mouse. *Cell Rep.* **10**, 292–305 (2015).
8. Gunaydin, L. A. et al. Natural neural projection dynamics underlying social behavior. *Cell* **157**, 1535–1551 (2014).
9. Graving, J. M. et al. DeepPoseKit, a software toolkit for fast and robust animal pose estimation using deep learning. *eLife* **8**, e47994 (2019).
10. Mathis, A. et al. DeepLabCut: markerless pose estimation of user-defined body parts with deep learning. *Nat. Neurosci.* **21**, 1281–1289 (2018).
11. Pereira, T. D. et al. Fast animal pose estimation using deep neural networks. *Nat. Methods* **16**, 117–125 (2019).
12. Geuther, B. Q. et al. Robust mouse tracking in complex environments using neural networks. *Commun. Biol.* **2**, 124 (2019).
13. Gris, K. V., Coutu, J.-P. & Gris, D. Supervised and unsupervised learning technology in the study of rodent behavior. *Front. Behav. Neurosci.* **11**, 141 (2017).
14. Schaefer, A. T. & Claridge-Chang, A. The surveillance state of behavioral automation. *Curr. Opin. Neurobiol.* **22**, 170–176 (2012).
15. Robie, A. A., Seagraves, K. M., Egnor, S. E. R. & Branson, K. Machine vision methods for analyzing social interactions. *J. Exp. Biol.* **220**, 25–34 (2017).
16. Vu, M.-A. T. et al. A shared vision for machine learning in neuroscience. *J. Neurosci.* **38**, 1601–1607 (2018).
17. Goodwin, N. L., Nilsson, S. R. O., Choong, J. J. & Golden, S. A. Toward the explainability, transparency, and universality of machine learning for behavioral classification in neuroscience. *Curr. Opin. Neurobiol.* **73**, 102544 (2022).
18. Newton, K. C. et al. Lateral line ablation by ototoxic compounds results in distinct rheotaxis profiles in larval zebrafish. *Commun. Biol.* **6**, 1–15 (2023).
19. Jernigan, C. M., Stafstrom, J. A., Zaba, N. C., Vogt, C. C. & Sheehan, M. J. Color is necessary for face discrimination in the Northern paper wasp, *Polistes fuscatus*. *Anim. Cogn.* **26**, 589–598 (2022).
20. Dahake, A. et al. Floral humidity as a signal – not a cue – in a nocturnal pollination system. Preprint at *bioRxiv* https://doi.org/10.1101/2022.04.27.489805 (2022).
21. Dawson, M. et al. Hypocretin/orexin neurons encode social discrimination and exhibit a sex-dependent necessity for social interaction. *Cell Rep.* **42**, 112815 (2023).
22. Baleisyte, A., Schneggenburger, R. & Kochubey, O. Stimulation of medial amygdala GABA neurons with kinetically different channelrhodopsins yields opposite behavioral outcomes. *Cell Rep.* **39**, 110850 (2022).
23. Cruz-Pereira, J. S. et al. Prebiotic supplementation modulates selective effects of stress on behavior and brain metabolome in aged mice. *Neurobiol. Stress* **21**, 100501 (2022).
24. Linders, L. E. et al. Stress-driven potentiation of lateral hypothalamic synapses onto ventral tegmental area dopamine neurons causes increased consumption of palatable food. *Nat. Commun.* **13**, 6898 (2022).
25. Slivicki, R. A. et al. Oral oxycodone self-administration leads to features of opioid misuse in male and female mice. *Addiction Biol.* **28**, e13253 (2023).
26. Miczek, K. A. et al. Excessive alcohol consumption after exposure to two types of chronic social stress: intermittent episodes vs. continuous exposure in C57BL/6J mice with a history of drinking. *Psychopharmacology (Berl.)* **239**, 3287–3296 (2022).
27. Cui, Q. et al. Striatal direct pathway targets Npas1^{+} pallidal neurons. *J. Neurosci.* **41**, 3966–3987 (2021).
28. Chen, J. et al. A MYT1L syndrome mouse model recapitulates patient phenotypes and reveals altered brain development due to disrupted neuronal maturation. *Neuron* **109**, 3775–3792 (2021).
29. Rigney, N., Zbib, A., de Vries, G. J. & Petrulis, A. Knockdown of sexually differentiated vasopressin expression in the bed nucleus of the stria terminalis reduces social and sexual behaviour in male, but not female, mice. *J. Neuroendocrinol.* **34**, e13083 (2021).
30. Winters, C. et al. Automated procedure to assess pup retrieval in laboratory mice. *Sci. Rep.* **12**, 1663 (2022).
31. Neira, S. et al. Chronic alcohol consumption alters home-cage behaviors and responses to ethologically relevant predator tasks in mice. *Alcohol Clin. Exp. Res.* **46**, 1616–1629 (2022).
32. Kwiatkowski, C. C. et al. Quantitative standardization of resident mouse behavior for studies of aggression and social defeat. *Neuropsychopharmacology* **46**, 1584–1593 (2021).
33. Yamaguchi, T. et al. Posterior amygdala regulates sexual and aggressive behaviors in male mice. *Nat. Neurosci.* **23**, 1111–1124 (2020).
34. Nygaard, K. R. et al. Extensive characterization of a Williams syndrome murine model shows Gtf2ird1-mediated rescue of select sensorimotor tasks, but no effect on enhanced social behavior. *Genes Brain Behav.* **22**, e12853 (2023).
35. Ojanen, S. et al. Interneuronal GluK1 kainate receptors control maturation of GABAergic transmission and network synchrony in the hippocampus. *Mol. Brain* **16**, 43 (2023).
36. Hon, O. J. et al. Serotonin modulates an inhibitory input to the central amygdala from the ventral periaqueductal gray. *Neuropsychopharmacology* **47**, 2194–2204 (2022).
37. Murphy, C. A. et al. Modeling features of addiction with an oral oxycodone self-administration paradigm. Preprint at *bioRxiv* https://doi.org/10.1101/2021.02.08.430180 (2021).
38. Neira, S. et al. Impact and role of hypothalamic corticotropin releasing hormone neurons in withdrawal from chronic alcohol consumption in female and male mice. *J. Neurosci.* **43**, 7657–7667 (2023).
39. Lapp, H. E., Salazar, M. G. & Champagne, F. A. Automated maternal behavior during early life in rodents (AMBER) pipeline. *Sci. Rep.* **13**, 18277 (2023).
40. Barnard, I. L. et al. High-THC cannabis smoke impairs incidental memory capacity in spontaneous tests of novelty preference for objects and odors in male rats. *eNeuro* **10**, ENEURO.0115-23.2023 (2023).
41. Ausra, J. et al. Wireless battery free fully implantable multimodal recording and neuromodulation tools for songbirds. *Nat. Commun.* **12**, 1968 (2021).
42. Friard, O. & Gamba, M. BORIS: a free, versatile open-source event-logging software for video/audio coding and live observations. *Methods Ecol. Evol.* **7**, 1325–1330 (2016).
43. Spink, A. J., Tegelenbosch, R. A. J., Buma, M. O. S. & Noldus, L. P. J. J. The EthoVision video tracking system—a tool for behavioral phenotyping of transgenic mice. *Physiol. Behav.* **73**, 731–744 (2001).
44. Lundberg, S. shap. https://github.com/shap/shap
45. Lauer, J. et al. Multi-animal pose estimation, identification and tracking with DeepLabCut. *Nat. Methods* **19**, 496–504 (2022).
46. Pereira, T. D. et al. SLEAP: a deep learning system for multi-animal pose tracking. *Nat Methods* **19**, 486–495 (2022).
47. Segalin, C. et al. The Mouse Action Recognition System (MARS) software pipeline for automated analysis of social behaviors in mice. *eLife* **10**, e63720 (2021).
48. Breiman, L. Random forests. *Mach. Learn.* **45**, 5–32 (2001).
49. Liaw, A. & Wiener, M. Classification and regression by randomForest. *R News* **2/3** https://journal.r-project.org/articles/RN-2002-022/RN-2002-022.pdf (2022).
50. Goodwin, N. L., Nilsson, S. R. O. & Golden, S. A. Rage against the machine: advancing the study of aggression ethology via machine learning. *Psychopharmacology* **237**, 2569–2588 (2020).

51. Lundberg, S. M. et al. From local explanations to global understanding with explainable AI for trees. *Nat. Mach. Intell.* **2**, 56–67 (2020).
52. Ribeiro, M. T., Singh, S., & Guestrin, C. 'Why should I trust you?': explaining the predictions of any classifier. Preprint at *arXiv* https://doi.org/10.48550/arXiv.1602.04938 (2016).
53. Sundararajan, M., Taly, A. & Yan, Q. Axiomatic attribution for deep networks. In *Proc. of the 34th International Conference on Machine Learning* 3319–3328 (MLR Press, 2017).
54. Hatwell, J., Gaber, M. M. & Azad, R. M. A. CHIRPS: explaining random forest classification. *Artif. Intell. Rev.* **53**, 5747–5788 (2020).
55. Lundberg, S. & Lee, S.-I. A unified approach to interpreting model predictions. Preprint at *arXiv* https://doi.org/10.48550/arXiv.1705.07874 (2017).
56. Verma, S., Dickerson, J. & Hines, K. Counterfactual explanations for machine learning: a review. Preprint at *arXiv* https://doi.org/10.48550/arXiv.2010.10596 (2020).
57. Rudin, C. Stop explaining black box machine learning models for high stakes decisions and use interpretable models instead. *Nat. Mach. Intell.* **1**, 206–215 (2019).
58. Takahashi, A. et al. Establishment of a repeated social defeat stress model in female mice. *Sci. Rep.* **7**, 12838 (2017).
59. Hashikawa, K. et al. Esr1^{+} cells in the ventromedial hypothalamus control female aggression. *Nat. Neurosci.* **20**, 1580–1590 (2017).
60. Newman, E. L. et al. Fighting females: neural and behavioral consequences of social defeat stress in female mice. *Biol. Psychiatry* **86**, 657–668 (2019).
61. Aubry, A. V. et al. Sex differences in appetitive and reactive aggression. *Neuropsychopharmacology* **47**, 1746–1754 (2022).
62. Golden, S. A., Covington, H. E., Berton, O. & Russo, S. J. A standardized protocol for repeated social defeat stress in mice. *Nat. Protoc.* **6**, 1183–1191 (2011).
63. Shemesh, Y. & Chen, A. A paradigm shift in translational psychiatry through rodent neuroethology. *Mol. Psychiatry* **28**, 993–1003 (2023).
64. Kabra, M., Robie, A. A., Rivera-Alba, M., Branson, S. & Branson, K. JAABA: interactive machine learning for automatic annotation of animal behavior. *Nat. Methods* **10**, 64–67 (2013).
65. Bordes, J. et al. Automatically annotated motion tracking identifies a distinct social behavioral profile following chronic social defeat stress. *Nat. Commun.* **14**, 4319 (2023).
66. Winters, C., Gorssen, W., Wöhr, M. & D'Hooge, R. BAMBI: a new method for automated assessment of bidirectional early-life interaction between maternal behavior and pup vocalization in mouse dam-pup dyads. *Front. Behav. Neurosci.* **17**, 1139254 (2023).
67. Lundberg, S. M., Erion, G. G. & Lee, S.-I. Consistent individualized feature attribution for tree ensembles. Preprint at *arXiv* https://doi.org/10.48550/arXiv.1802.03888 (2019).
68. Covert, I. C., Lundberg, S. & Lee, S.-I. Explaining by removing: a unified framework for model explanation. *J. Mach. Learn. Res.* **22**, 1–90 (2021).
69. Lorbach, M., Poppe, R. & Veltkamp, R. C. Interactive rodent behavior annotation in video using active learning. *Multimed. Tools Appl.* **78**, 19787–19806 (2019).
70. Tillmann, J. F., Hsu, A. I., Schwarz, M. K. & Yttri, E. A. A-SOiD, an active-learning platform for expert-guided, data-efficient discovery of behavior. *Nat. Methods* **21**, 703–711 (2024).
71. Whiteway, M. R. et al. Partitioning variability in animal behavioral videos using semi-supervised variational autoencoders. *PLoS Comput. Biol.* **17**, e1009439 (2021).
72. Sun, J. J. et al. Task Programming: Learning Data Efficient Behavior Representations. In *Proc IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.* 2875–2884 (2021).
73. MABe 2022. Multi-agent behavior: representation, modeling, measurement, and applications. https://sites.google.com/view/mabe22/home
74. Sun, J. J. et al. The multi-agent behavior dataset: mouse dyadic social interactions. Preprint at *arXiv* https://doi.org/10.48550/arXiv.2104.02710 (2021).
75. OpenBehavior. About the OpenBehavior Project and the open source movement. https://edspace.american.edu/openbehavior/
76. Mouse Phenome Database. https://phenome.jax.org/about
77. Kapoor, S. & Narayanan, A. Leakage and the reproducibility crisis in machine-learning-based science. *Patterns (N Y)* **4**, 100804 (2023).
78. Dankert, H., Wang, L., Hoopfer, E. D., Anderson, D. J. & Perona, P. Automated monitoring and analysis of social behavior in *Drosophila*. *Nat. Methods* **6**, 297–303 (2009).
79. de Chaumont, F. et al. Real-time analysis of the behaviour of groups of mice via a depth-sensing camera and machine learning. *Nat. Biomed. Eng.* **3**, 930–942 (2019).
80. Giancardo, L. et al. Automatic visual tracking and social behaviour analysis with multiple mice. *PLoS ONE* **8**, e74557 (2013).
81. Hong, W. et al. Automated measurement of mouse social behaviors using depth sensing, video tracking, and machine learning. *Proc. Natl Acad. Sci. USA* **112**, E5351–E5360 (2015).
82. Goodwin, N. L. et al. Simple behavioral analysis (SimBA) as a platform for explainable machine learning in behavioral neuroscience. *Nat. Neurosci.* (in the press).
83. Bohnslav, J. P. et al. DeepEthogram, a machine learning pipeline for supervised behavior classification from raw pixels. *eLife* **10**, e63377 (2021).
84. Gerós, A., Magalhães, A. & Aguiar, P. Improved 3D tracking and automated classification of rodents' behavioral activity using depth-sensing cameras. *Behav. Res.* **52**, 2156–2167 (2020).
85. Harris, C., Finn, K. R., Kieseler, M.-L., Maechler, M. R. & Tse, P. U. DeepAction: a MATLAB toolbox for automated classification of animal behavior in video. *Sci. Rep.* **13**, 2688 (2023).
86. Hu, Y. et al. *LabGym*: quantification of user-defined animal behaviors using learning-based holistic assessment. *Cell Rep. Methods* **3**, 100415 (2023).
87. Marks, M. et al. Deep-learning based identification, tracking, pose estimation, and behavior classification of interacting primates and mice in complex environments. *Nat. Mach. Intell.* **4**, 331–340 (2022).
88. Branson, K., Robie, A. A., Bender, J., Perona, P. & Dickinson, M. H. High-throughput ethomics in large groups of *Drosophila*. *Nat. Methods* **6**, 451–457 (2009).
89. Jia, Y. et al. Selfee, self-supervised features extraction of animal behaviors. *eLife* **11**, e76218 (2022).
90. Berman, G. J., Choi, D. M., Bialek, W. & Shaevitz, J. W. Mapping the stereotyped behaviour of freely moving fruit flies. *J. R. Soc. Interface* **11**, 20140672 (2014).
91. Arakawa, T. et al. Automated estimation of mouse social behaviors based on a hidden Markov model. In *Hidden Markov Models: Methods and Protocols* (eds Westhead, D. R. & Vijayabaskar, M. S.) 185–197 (Humana Press, 2017).
92. Chen, Z. et al. AlphaTracker: a multi-animal tracking and behavioral analysis tool. *Front. Behav. Neurosci.* **17**, 1111908 (2023).
93. Huang, K. et al. A hierarchical 3D-motion learning framework for animal spontaneous behavior mapping. *Nat. Commun.* **12**, 2784 (2021).

94. Luxem, K. et al. Identifying behavioral structure from deep variational embeddings of animal motion. *Commun. Biol.* **5**, 1267 (2022).
95. Nandi, A., Virmani, G., Barve, A. & Marathe, S. DBscorer: an open-source software for automated accurate analysis of rodent behavior in forced swim test and tail suspension test. *eNeuro* **8**, ENEURO.0305-21.2021 (2021).
96. Gabriel, C. J. et al. BehaviorDEPOT is a simple, flexible tool for automated behavioral detection based on markerless pose tracking. *eLife* **11**, e74314 (2022).

[1]Department of Biological Structure, University of Washington, Seattle, WA, USA. [2]Graduate Program in Neuroscience, University of Washington, Seattle, WA, USA. [3]Center of Excellence in Neurobiology of Addiction, Pain and Emotion (NAPE), University of Washington, Seattle, WA, USA. [4]Department of Electrical and Computer Engineering, University of Washington, Seattle, WA, USA. [5]New York University Neuroscience Institute, New York, NY, USA. [6]Department of Psychiatry, Harvard Medical School McLean Hospital, Belmont, MA, USA. [7]Department of Psychology, Tufts University, Medford, MA, USA. [8]Department of Integrative Physiology and Neuroscience, Washington State University, Pullman, WA, USA. [9]Graduate Program in Neuroscience, Washington State University, Pullman, WA, USA. [10]Stanford University School of Medicine, Stanford, CA, USA. [11]Department of Psychiatry and Behavioral Sciences, Stanford University, Stanford, CA, USA. [12]Department of Anesthesiology and Pain Medicine, University of Washington, Seattle, WA, USA. [13]These authors contributed equally: Simon R. O. Nilsson, Sam A. Golden. ✉e-mail: sronilsson@gmail.com; sagolden@uw.edu

Methods

Animals

Male CD-1 (strain no. 022, Charles River Laboratories (CRL)) or female CFW (strain no. 024, CRL) white coat colored 12-week-old mice were used as residents. Female CFW residents were pair housed with 10-week-old castrated CFW male mice for the duration of the study to elicit aggression[60,61], whereas male CD-1 residents were singly housed. Sex-matched, black coat colored C57BL6/J (C57; strain no. 000664, The Jackson Laboratory) mice were used as intruders for all aggression assays. C57 females were more than 10 weeks old, and males were more than 8 weeks old. We chose to use differently coated social partners to facilitate subject identification, which is improved by using social pairs with different coat colors. We gave all mice free access to standard food chow and water in all experiments. We housed all mice with enrichment (cotton padding) in standard Allentown clear polycarbonate cages covered with stainless steel wire lids at least 1 week before experiments, and we maintained them on a reverse 12-h light/dark cycle (light off at 9:00).

Male Long-Evans rats (Simonsen Laboratories, 120–140 d old) were used as residents. Residents were housed in isolation in standard rat cages for the duration of the study. Male Sprague-Dawley rats (bred in-house, 60–80 d old) were used as intruders. Intruders were pair housed in standard rat cages. We maintained all animals on a reverse 12-h light/dark cycle (light off at 7:00) and provided ad libitum access to food and water. No statistical methods were used to pre-determine sample sizes, but our sample sizes are similar to those reported in previous publications[60,62,97]. Animals were randomly assigned to social defeat and context groups.

All animals were housed in humidity-controlled rooms. All experiments were performed in accordance with the Guide for the Care and Use of Laboratory Animals under protocols approved by the local animal care and use committee at each institution (University of Washington, Tufts University, Stanford University, Columbia University, Washington State University and the Icahn School of Medicine at Mount Sinai).

Code base

SimBA is maintained on GitHub (https://github.com/sgoldenlab/simba), pip (https://pypi.org/project/Simba-UW-tf-dev/), Read the Docs (https://simba-uw-tf-dev.readthedocs.io/) and Gitter (https://app.gitter.im/#/room/#SimBA-Resource_community). SimBA was originally conceptualized by S.R.O.N., N.G. and S.A.G. Version 1 was released in 2020 with code written by S.N. and J.J.C., with minor contributions from N.G., A.I. and S.H. Since then, the SimBA version 2 codebase, documentation and support forums are written, maintained and developed independently by S.R.O.N., with important non-coding support from N.G. The authors sincerely thank the collaborative and friendly open-source community for their continued expert feedback and suggestions.

Behavioral protocols

We used male[62] and female[60,61] CSDS procedures as previously published. In brief, we recorded dyadic encounters between male or female mice in clear polycarbonate cages (cage size: 28 × 19 × 12 cm) divided in half by a clear acrylic barrier. Aggressive CD-1 or CFW resident animals were housed on one side of the barrier where they encountered an unfamiliar sex-matched C57 intruder for 5 min (n = 11 female residents, n = 10 intruders) or 10 min (n = 21 male residents, n = 21 intruders). Female CFW mice were rendered aggressive through cohabitation with castrated males; castrated males were temporarily single housed during daily 5-min female defeat sessions. We analyzed 5 d of data per aggressive mouse.

For RI procedures, we recorded dyadic encounters between male mice (n = 24 residents, n = 24 group-housed intruders) in standard shoebox home cages (17.8 × 28.6 cm). C57 intruder animals were introduced into the center of the CD-1 resident home cage for 10 min and allowed to freely interact. Animals were tested once a day for three consecutive days. Rat RI assays were recorded between same-sex conspecifics in clear polycarbonate cages with fresh bedding (cage size: 33 × 46 × 19 cm). Residents were tested three times 4–8 weeks apart with novel intruders. Assays lasted for up to 10 min.

Video recordings

All recordings were made overhead. Male mice were recorded at 30–80 fps with USB 3.0 cameras (acA2040-120uc Basler ace, Basler) using fixed focal length lenses (Edmund Optics, 16 mm/F1.4) at variable resolutions (W: 1,000–1,200 pixels, H: 1,255–2,056 pixels) using pylon camera software (Basler). Male rats and female mice were recorded at 30–60 fps and 1,280 × 720 resolution using a Logitech C922 camera. All recorded videos used to build behavioral classifiers were re-sampled in SimBA to 30 fps before being used to create machine learning classifiers. The Caltech Resident-Intruder Mouse 13 (CRIM13)[98] dataset was recorded at 25 fps and 640 × 480 resolution.

Video processing

Video recordings were pre-processed using tools available in SimBA. We shortened, cropped and saved videos and frames in RGB, gray-scale and contrast-limited adaptive histogram equalization (CLAHE) enhanced formats at variable frame rates. CLAHE video conversion can improve image quality and pose estimation in non-optimal recording conditions. The CRIM13 dataset was recorded in SEQ file format and converted to MP4 format using in-built functions in SimBA (provided by Xiaoyu Tong, Lin laboratory, New York University). We used SimBA to extract frames within specific time periods, concatenate videos, downsample video resolutions and generate gifs and videos. An exhaustive list of SimBA video pre-processing tools and tutorials is available on the SimBA GitHub repository and in Supplementary Note 2—SimBA software manual.

Pose estimation

We created pose estimation models in DeepLabCut (version 1.0)[10] for rats, the CRIM dataset and our experimental mouse videos. SimBA currently supports pose estimation import from regular and maDeepLabCut[10,45], SLEAP[46], BENTO[72], DANNCE 3D[99], DeepPoseKit[9] and Animal Part Tracker[100]. In rat and CRIM datasets, we tracked eight points per animal, consisting of nose, left and right ears, left and right sides, back, tail base and tail end. For our mouse model, we tracked seven points per animal (same points excluding tail end). We labeled frames from a diverse set of videos to create training sets for ResNet-50-based neural networks for all models. The pose estimation data were imported to SimBA for further analysis. All models are available on the SimBA Open Science Framework (OSF) repository. Furthermore, SimBA includes methods for cross-video animal tracking swapping and pose scheme body part dropping to facilitate merging pose estimation datasets (Supplementary Note 2—SimBA software manual).

Classifier creation

Our original classifier set[82] consisted of 10 mouse classifiers (attack, pursuit, lateral threat, anogenital sniffing, allogrooming normal, allogrooming vigorous, mounting, scramble, flee and upright submissive); seven rat classifiers (attack, anogenital sniffing, lateral threat, approach, boxing, avoidance and submission); and 11 mouse classifiers from the open-source CRIM13 (ref. 98) dataset (approach, attack, chase, circle, copulation, drink, eat, sniff, up, clean and walk away). All rat and CRIM13 data presented in this manuscript use these original classifiers. The five University of Washington mouse classifiers (attack, anogenital sniffing, pursuit, escape and defensive) are optimized iterations of a subset of the original 10 mouse classifiers. All datasets are available, with detailed information on classifier creation, on the SimBA OSF repository. Importantly, all frames included in the random forest training sets need definite labels as positive or negative for the

behavior of interest. Incorrectly labeled events negatively impact model performance. As such, we created strict operational definitions for our behaviors (Supplementary Figs. 1 and 2).

We used SimBA to create random forest classifiers using scikit-learn with one classifier per behavior of interest (Supplementary Note 2—SimBA software manual). Random forest classifiers accept a range of hyperparameters that specify how decision trees are generated. Hyperparameters for our provided classifiers consist of the following: criterion = entropy, max features = square root, minimum sample leaf = 1 and number of estimators = 2,000. Random undersampling was used in cases of major class imbalances (Extended Data Fig. 2). All classifiers were iteratively trained and optimized (Fig. 2c). SimBA accepts a range of different random forest hyperparameter settings and sampling methods, but users unfamiliar with the available parameters can import recommended or previously successful settings (Supplementary Note 2—SimBA software manual) based on classifiers of behaviors with similar frequency and salience.

Classifier performance

We used SimBA to generate classifier learning curves (Fig. 2a, left). In these learning curves, we evaluated F1 scores after performing five-fold cross-validations using 1%, 25%, 50%, 75% and 100% of the shuffled datasets to predict the classified behaviors on 20% of the datasets. Learning curves indicate how inclusion of further logged behavioral events affect classifier performance. Furthermore, we used SimBA to generate precision-recall curves (Fig. 2a, right), which inform the balance between sensitivity and specificity at varying classifier discrimination thresholds. F1 scores are a harmonic mean of precision and recall: F1 = 2 × (precision × recall / (precision + recall)), where precision quantifies the number of positive class predictions that actually belong to the positive class, and recall quantifies the number of positive class predictions made out of all positive examples in the dataset.

SHAP

The Shapley value contribution[68] of a specific feature $\varnothing_i$ toward the final prediction generated by model f for a data instance with features x is computed as follows. For a given data instance, x, for all possible subsets $S \subseteq F$, where F represents the full set of features used in the model, the model is trained both with the feature of interest present and without it present, denoted $f_{S\cup i}$ and f_S, respectively. Predictions generated from the model excluding the feature are subtracted from the predictions generated from the model including the feature $f(x_{S\cup i}) - f(x_S)$. The difference in predicted values is multiplied by the number of possible permutations of the set, multiplied by the number of permutations of the remaining features, divided by the total possible number of permutations of features, where $|F|$ is the cardinality of the feature set F, ultimately generating a weighted average. The weighted averages of each set contribution are summed to generate the total contribution of feature i. Thus, the contribution of feature i to the prediction is:

$$\phi_i(f,x) = \sum_{S\subseteq F\setminus i} \frac{|S|!(|F|-|S|-1)!}{|F|!}\left(f_{S\cup i}(x_{S\cup i}) - fs(x_S)\right)$$

Explainability and classifier comparisons

We used SHAP[55] and TreeSHAP[67] to evaluate how feature values impact classification probabilities (Figs. 3–6). Owing to the Shapley additivity axiom, we can collapse Shapley value contributions of features that measure similar, general and often colinear characteristics of the behavior of interest into interpretable physical categories while maintaining consistency, accuracy and biological relevance. Hence, to aid interpretability[17], we collapsed the features into seven behaviorally defined feature categories that measure general characteristics of social interactions (that is, animal distances, resident and intruder movement, intruder movement, resident movement, resident shape, intruder shape and resident and intruder shape). Each of the seven feature categories contained six temporal windows that represent different frame sampling frequencies (single frame to 500 ms). The list of features encompassed by each collapsed feature category is modifiable for custom use cases in SimBA (Supplementary Note 2—SimBA software manual, 'Mixins'). We created the general structure of the feature categories to ensure compatibility with alternative and novel feature sets targeting similar behaviors.

The use of these categories ensures that classifiers targeting the same behavior, but using different feature sets, can be directly compared. A common reason that classifiers use different feature sets is a consequence of the initial pose estimation scheme selected by experimenters, which influences the selection of features. For example, two research groups interested in the same behavior may use pose estimation schemes that identify five versus eight body parts due to their experimental needs; regardless, by binning their classification features within the same biologically determined classes and sampling frequencies, the use of Shapley values allows for direct comparisons between their classifiers.

SimBA was used for evaluating feature SHAP values in behavioral classifiers (Supplementary Note 1—SimBA software manual). The classifiers for cross-site SHAP comparisons (Fig. 3) were generated based on annotations from four different laboratories and showed similar performance for classifying attack behavior within their respective datasets (precision > 0.937, recall > 80.0, F1 > 86.3, attack present > 4,437). For comparisons between rat and mouse attack classifiers (Fig. 4), a random sample of 4,000 attack and 4,000 non-attack frames from each site was analyzed using SHAP (total: 32,000 frames). Each of the five updated mouse classifiers (Fig. 2) was similarly evaluated on 10,000 (or greatest amount if fewer than 10,000 available) positive frames per behavior (Figs. 5 and 6). From the supervised analysis of the biological datasets, we extracted 1,000 positive frames before Markov model[101] smoothing in SimBA and evaluated these for SHAP values (Figs. 5 and 6).

Behavioral analysis

For the CSDS and RI datasets, we imported DeepLabCut pose information into SimBA. We used the SimBA GUI to calibrate video scales (pixels per millimeter). Outlier correction was performed using heuristic rules as documented in the SimBA API (movement criterion: 0.7, location criterion: 1.5). We extracted 498 features and analyzed videos with five classifiers (resident: attack, anogenital sniffing, pursuit. intruder: escape, defensive) using parameters outlined in Fig. 2. Multiple behaviors could be present in a single frame (that is, attack and defensive behavior). We used a Kleinberg filter[52] (sigma: 0.3; kappa: 2; hierarchy: 2 for anogenital and attack, 3 for defensive and escape and 4 for pursuit) with the hierarchical search function for all classifiers as documented in the SimBA API. The Kleinberg algorithm uses an infinite Markov chain[53] to delineate hierarchical temporal sequences, or 'bursts', where classified events are more or less likely. For this, we modified a version of the pyburst package available in SimBA (Supplementary Note 2—SimBA software manual). We analyzed the final machine learning results by calculating the total event duration, number of bouts, mean bout duration and mean bout interval per video.

Statistical analysis

We visualized and calculated descriptive statistics of attack, anogenital sniffing, pursuit, escape and defensive behaviors for three groups of mice using SimBA in-built functions (Figs. 5 and 6). For each classified behavior, we used SimBA to calculate total behavior duration, bouts per session, mean bout duration and mean bout interval. For sex comparisons, the first 5 min of the male and female CSDS assays were used across 5 d. For environmental context comparisons (CSDS versus RI), the full 10 min of RI and CSDS assays were used from the first 3 d of testing. Separate attack classifiers were built and used to analyze

behavior for rats and each of the mouse laboratories participating in the SHAP consortium dataset. No datapoints were excluded from analysis. Data distribution was assumed to be normal, but this was not formally tested. Data collection and analysis were not performed blinded to the conditions of the experiments. Data were analyzed in GraphPad Prism (version 8.0.1) via *t*-tests, two-way ANOVAs or linear mixed models as appropriate, and Bonferroni correction was used for multiple comparisons ($\alpha = 0.05$) ($^{*}P < 0.05$, $^{**}P \leq 0.01$ and $^{***}P \leq 0.001$ in all figures). Statistical comparisons are described in Supplementary Note 1—detailed statistics.

Reporting summary

Further information on research design is available in the Nature Portfolio Reporting Summary linked to this article.

Data availability

The data used in this study are available as a branch of the SimBA GitHub repository: https://github.com/sgoldenlab/simba/tree/simba-data/data.

The CRIM13 database is publicly available at https://paperswithcode.com/dataset/crim13. Source data are provided with this paper.

Code availability

Code and documentation are available at https://github.com/sgoldenlab/simba. SimBA was built with Python version 3.6.

References

97. Golden, S. A. et al. Epigenetic regulation of RAC1 induces synaptic remodeling in stress disorders and depression. *Nat. Med.* **19**, 337–344 (2013).
98. Burgos-Artizzu, X. P., Dollar, P., Lin D., Anderson, D. J. & Perona, P. CRIM13 (Caltech Resident-Intruder Mouse 13) (1.0). CaltechDATA. https://doi.org/10.22002/D1.1892 (2021).
99. Karashchuk, P., Tuthill, J. C. & Brunton, B. W. The DANNCE of the rats: a new toolkit for 3D tracking of animal behavior. *Nat. Methods* **18**, 460–462 (2021).
100. Branson, K. APT. https://github.com/kristinbranson/APT
101. Lee, W., Fu, J., Bouwman, N., Farago, P. & Curley, J. P. Temporal microstructure of dyadic social behavior during relationship formation in mice. *PLoS ONE* **14**, e0220596 (2019).

Acknowledgements

This research was supported by National Institute on Drug Abuse (NIDA) R00DA045662 (S.A.G.), R01DA059374 (S.A.G.) and P30DA048736 (M.H. and S.A.G.); NARSAD Young Investigator Award 27082 (S.A.G.); National Institute of Mental Health 1F31MH125587 (N.L.G.), F31AA025827 (E.L.N.) and F32MH125634 (E.L.N.); National Institute of General Medical Sciences R35GM146751 (M.H.); National Institutes of Health K08MH123791 (N.E.); the Burroughs Wellcome Fund Career Award for Medical Scientists (N.E.); the Simons Foundation Bridge to Independence Award (N.E.); and the Washington Research Foundation Postdoctoral Fellowship (E.R.S.). We thank V. Tsai, R. Vrooman and C. Xu for their skillful technical contributions. We thank B. Bentzley and D. Lin for contributing to aggression consortium data. S.R.O.N. has continued development and maintenance of SimBA independent of other funding sources. Figures were created with BioRender.

Author contributions

Conceptualization: S.A.G., S.R.O.N. and N.L.G. Data curation: K.P., L.B., A.I., Y.Y.Z., E.S., X.T., E.L.N., K.M., H.R.W., R.J.M., Z.C.N., N.E., M.H., S.R.O.N., N.L.G. and S.A.G. Methodology: S.R.O.N. and N.L.G. Formal analysis: S.R.O.N., N.L.G. and S.A.G. Visualization: N.L.G. and S.R.O.N. Writing—original draft: S.A.G., S.R.O.N. and N.L.G. Funding acquisition: S.A.G., S.R.O.N., N.L.G., M.H., E.L.N., N.E. and E.R.S. Writing—reviewing and editing: S.A.G., S.R.O.N., N.L.G., K.P., L.B., A.I., Y.Y.Z., E.S., X.T., E.L.N., K.M., H.R.W., R.J.M., Z.C.N., N.E. and M.H. Software: S.R.O.N., J.J.C. and S.H. Supervision: S.A.G. Project administration: S.R.O.N. and N.L.G.

Competing interests

The authors declare no competing interests.

Additional information

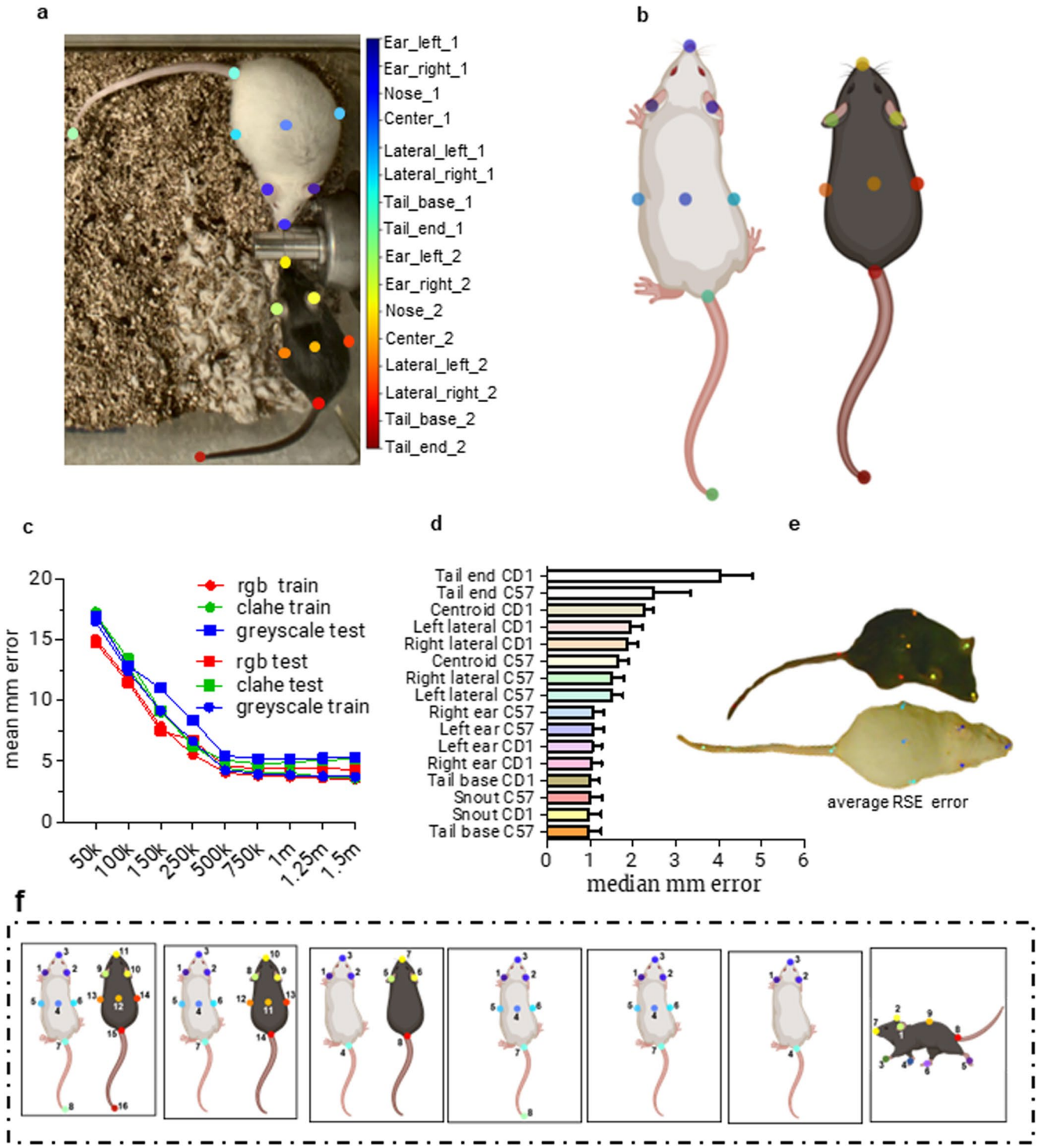
Extended data is available for this paper at https://doi.org/10.1038/s41593-024-01649-9.

Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s41593-024-01649-9.

Correspondence and requests for materials should be addressed to Simon R. O. Nilsson or Sam A. Golden.

Peer review information *Nature Neuroscience* thanks Joshua Shaevitz and the other, anonymous, reviewer(s) for their contribution to the peer review of this work.

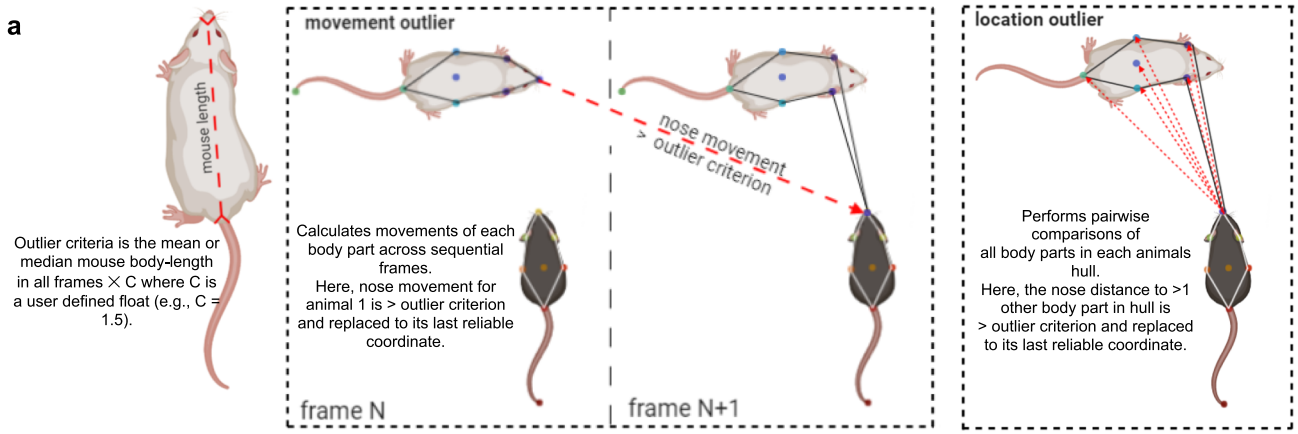
https://doi.org/10.1038/s41593-024-01649-9



Extended Data Fig. 1 | Example of DeepLabCut pose-estimation model for mouse resident-intruder behavior. These data were calculated from 12,686 images from n = 101 mice. (**a**) The 16 body-parts labeled. (**b**) Schematic depiction of the location of each of the 16 body part labels. (**c**) Evaluations of three models (rgb, clahe, greyscale) using the DeepLabCut evaluation tool. Pixel distances were converted to millimeter by using the lowest resolution images in the dataset (1000x1544px; 4.6px/millimeter). (**d**) Median millimeter error per body part. (**e**) Image representing the relative standard error (RSE) of the median millimeter error across all test images. The labelled images and DeepLabCut generated weights are available to download on the Open Science Framework, osf.io/mutws. (**f**) SimBA supports a range of alternative body-part settings for single animals and dyadic protocols through the File-> Create Project menu. Note: tail end tracking performance was insufficient for a tail rattle classifier, and the tail end body parts were dropped for all analysis in the main figures. Data are presented as mean +/− SEM.

https://doi.org/10.1038/s41593-024-01649-9

a



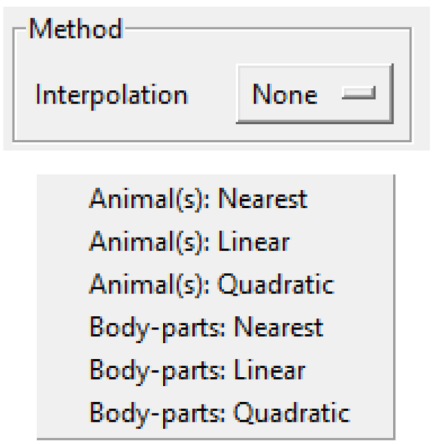
b

Body part	# frames	% of total
C57 tail end	24988	11.19
CD1 tail end	12261	5.49
CD1 right ear	2802	1.25
C57 tail base	1993	0.89
CD1 tail base	1423	0.64
C57 right ear	1141	0.51
CD1 centroid	268	0.12
C57 centroid	219	0.1
C57 left ear	155	0.07
CD1 left ear	141	0.06
CD1 lateral left	79	0.03
CD1 lateral right	40	0.02
C57 lateral left	53	0.02
C57 lateral right	51	0.02
C57 nose	35	0.02
CD1 nose	22	0.01

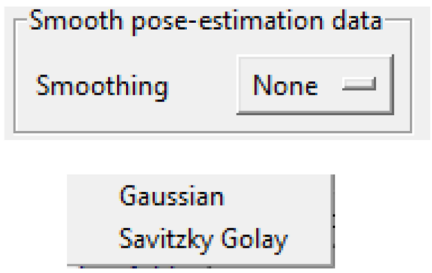
c

Body part	# frames	% of total
C57 tail base	3251	1.45
CD1 nose	2876	1.29
CD1 tail base	1624	0.73
CD1 left ear	625	0.28
CD1 right ear	565	0.25
CD1 lateral left	394	0.18
CD1 lateral right	375	0.17
CD1 centroid	350	0.16
C57 nose	360	0.16
C57 left ear	115	0.05
C57 right ear	88	0.04
C57 centroid	12	0.01
C57 lateral left	23	0.01
C57 lateral right	32	0.01

d



e



Extended Data Fig. 2 | See next page for caption.

https://doi.org/10.1038/s41593-024-01649-9

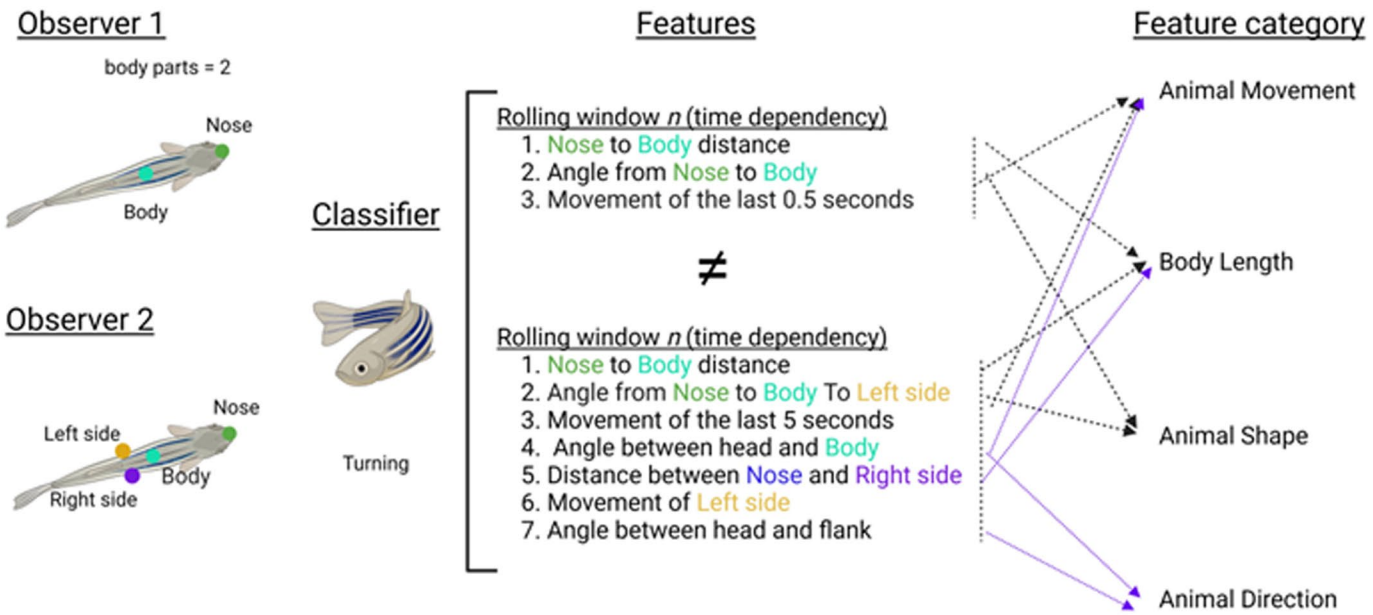
Extended Data Fig. 2 | SimBA outlier correction options. (**a**) SimBA calculates the mean or median distance between two user-defined body-parts across the frames of each video. We set the user-defined body-parts to be the nose and the tail-base of each animal. The user also defines a movement criterion value, and a location criterion value. We set the movement criterion to 0.7, and location criterion to 1.5. Two different outlier criteria are then calculated by SimBA. These criteria are the mean length between the two user-defined body parts in all frames of the video, multiplied by the either user-defined movement criterion value or location criterion value. SimBA corrects movement outliers prior to correcting location outliers. (**b**) Schematic representations of a pose-estimation body-part 'movement outlier' (top) and a 'location outlier' (bottom). A body-part violates the movement criterion when the movement of the body-part across sequential frames is greater than the movement outlier criterion. A body-part violates the location criteria when its distance to more than one other body-part in the animals' hull (except the tail-end) is greater than the location outlier criterion. Any body part that violates either the movement or location criterion is corrected by placing the body-part at its last reliable coordinate. (**c**) The ratio of body-part movements (top) and body-part locations (bottom) detected as outliers and corrected by SimBA in the RGB-format mouse resident-intruder data-set. For the outlier corrected in rat and the CRIM13 datasets, see the SimBA GitHub repository. We also offer (**d**) interpolation options for frames with missing body parts and (**e**) smoothing options to reduce frame-to-frame jitter.

https://doi.org/10.1038/s41593-024-01649-9

	classifier	# annotated frames	# annotated videos	annotations: % present	annotations: present (hh:mm:ss)	annotations: absent (hh:mm:ss)	under sample ratio	test set: frames present	test set: frames absent	optimal threshold
mouse resident-intruder	allogroom normal	334680	26	1.49	00:02:46	03:03:10	19	983	65953	0.48
	allogroom vigorous	334680	26	0.73	00:01:21	03:04:35	0	494	66442	0.41
	attack	203841	37	4.93	00:05:35	01:47:40	9	1920	38849	0.56
	lateral threat	168738	21	1.49	00:01:24	01:32:21	18	503	33245	0.52
	mounting	470294	31	2.88	00:07:31	04:13:45	4	2639	91420	0.72
	pursuit	103538	33	2.02	00:01:10	00:56:22	20	419	20289	0.43
	upright submissive	272304	20	1.13	00:01:43	02:29:34	16	577	53884	0.54
	flee	395852	22	1.08	00:02:23	03:37:33	28	877	78294	0.60
	scramble	323852	18	2.16	00:03:53	02:56:02	3.9	1347	63424	0.79
	anogenital sniffing	103538	33	6.57	00:03:47	00:53:45	8	1268	19440	0.55
rat resident-intruder	attack	136830	12	16.40	00:12:28	01:03:33	0	4437	32119	0.55
	lateral threat	136830	12	9.09	00:06:55	01:09:06	0	9943	26613	0.55
	anogenital sniffing	136830	12	7.50	00:05:42	01:10:19	7.5	2095	22713	0.45
	submission	136830	12	11.27	00:08:34	01:07:27	5	3074	33482	0.59
	boxing	136830	12	4.50	00:03:25	01:12:36	0	1306	23502	0.47
	approach	136830	12	4.23	00:03:13	01:12:48	8.5	1053	23755	0.50
	avoidance	136830	12	2.48	00:01:53	01:14:08	10	686	24122	0.48
	approach	757350	64	4.13	00:20:51	18:01:39	12	6190	145280	0.44
CRIM13 resident-intruder	attack	757350	64	4.10	00:20:42	18:01:48	16	6266	145204	0.40
	chase	757350	64	0.49	00:02:28	18:20:02	46	769	150701	0.58
	circle	757350	64	1.32	00:06:40	18:15:50	22	1982	149488	0.54
	clean	757350	64	7.99	00:40:20	17:42:10	8	12095	139375	0.53
	copulation	757350	64	4.62	00:23:20	17:59:10	12	7080	144390	0.50
	drink	757350	64	0.36	00:01:49	18:20:41	10	570	150900	0.54
	eat	757350	64	2.54	00:12:49	18:09:41	12	3824	147646	0.43
	sniff	757350	64	14.94	01:15:26	17:07:04	0	22757	128713	0.54
	up	757350	64	3.71	00:18:44	18:03:46	6	5575	145895	0.50
	walk away	757350	64	3.89	00:19:38	18:02:52	8	5897	145573	0.50

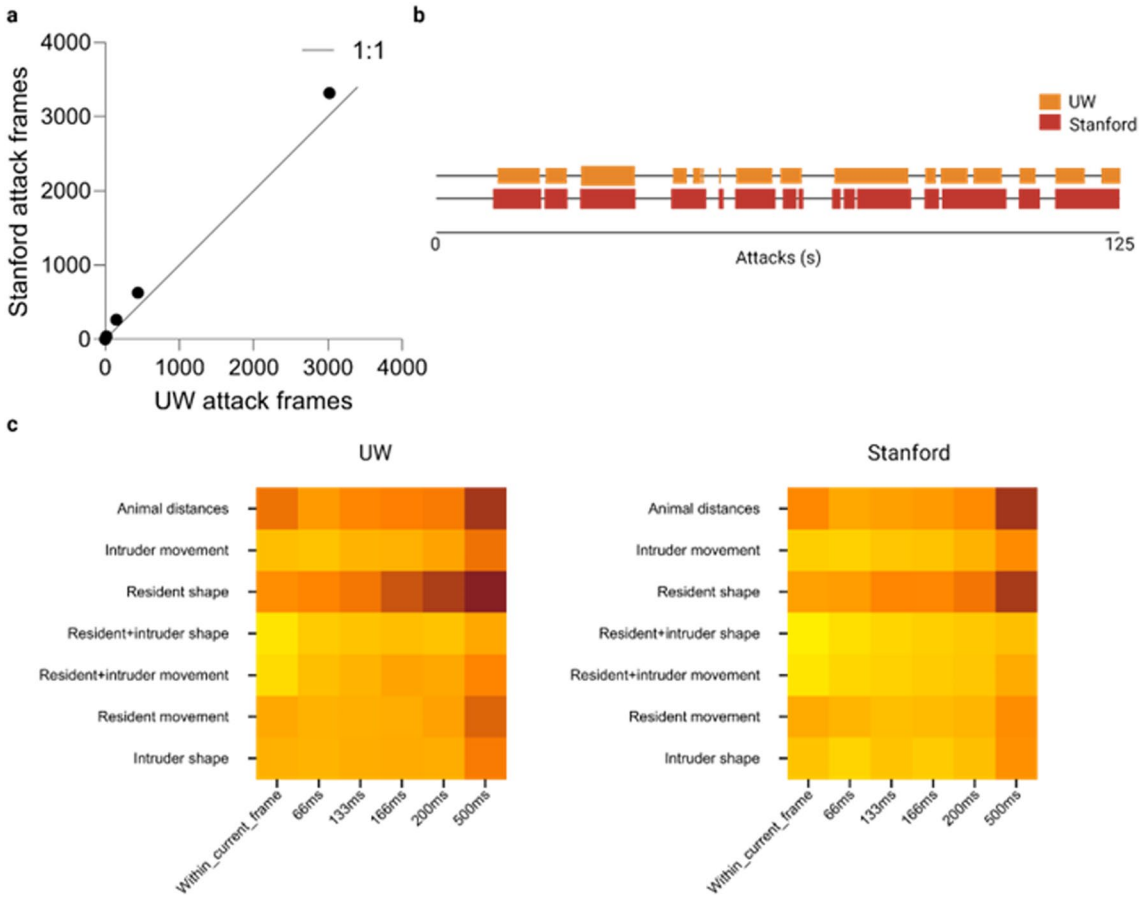
Extended Data Fig. 3 | Training set information for original mouse, rat, and CRIM13 classifiers. Training set information for mouse, rat, and CRIM13 mouse resident intruder behavioral classifiers.

https://doi.org/10.1038/s41593-024-01649-9



Extended Data Fig. 4 | Feature binning for SHAP calculations. Classifiers for the same behavior using different pose estimation schemes will have different feature lists, but can be directly compared via feature binning through the SHAP additivity axiom.

https://doi.org/10.1038/s41593-024-01649-9



Extended Data Fig. 5 | UW versus Stanford scoring and SHAP scores. UW and Stanford manual scoring of the same dataset for attack behavior. (**a**) Manual annotations (n = 9 videos) were highly correlated (R2 = 0.998). (**b**) Gantt plot of UW versus Stanford scores for a high-attack video. (**c**) SHAP scores for UW positive or Stanford positive attack frames. UW scores rely more on longer rolling windows of behavior than Stanford does.

https://doi.org/10.1038/s41593-024-01649-9

Attack SHAP (Positive frames)	0ms	66ms	133ms	166ms	200ms	500ms
Animal distances	0.001397	0.001544	0.002811	0.003851	0.004815	0.015507
Intruder movement	0.008883	0.001855	0.003253	0.004434	0.00524	0.01376
Resident + intruder movement	0.002091	0.006047	0.006217	0.00688	0.007653	0.010801
Resident movement	0.000943	0.004572	0.008186	0.010696	0.013132	0.03583
Intruder shape	0.08918	0.012769	0.012295	0.012219	0.013057	0.015059
Resident + intruder shape	0.051291	0.018518	0.018442	0.020436	0.020672	0.028871
Resident shape	0.032857	0.025453	0.028869	0.033255	0.033684	0.051612

Extended Data Fig. 6 | SHAP values for rat attack classifier. SHAP values across feature bins and rolling windows for rat attack classifier.

https://doi.org/10.1038/s41593-024-01649-9

Attack SHAP (Positive frames)	0ms	66ms	133ms	166ms	200ms	500ms
Animal distances	0.0405	0.0377	0.0343	0.0370	0.0331	0.0194
Intruder movement	0.0173	0.0148	0.0326	0.0378	0.0482	0.0939
Resident + intruder movement	0.0011	0.0074	0.0179	0.0291	0.0365	0.0968
Resident movement	0.0030	0.0039	0.0059	0.0066	0.0076	0.0184
Intruder shape	0.0061	0.0063	0.0081	0.0082	0.0088	0.0104
Resident + intruder shape	0.0006	0.0011	0.0013	0.0015	0.0016	0.0025
Resident shape	0.0026	0.0036	0.0042	0.0044	0.0049	0.0066

Pursuit SHAP (Positive frames)	0ms	66ms	133ms	166ms	200ms	500ms
Animal distances	0.0704	0.0187	0.0192	0.0184	0.0187	0.0195
Intruder movement	0.0042	0.0048	0.0136	0.0238	0.0207	0.0721
Resident + intruder movement	0.0011	0.0074	0.0197	0.0274	0.0390	0.0781
Resident movement	0.0056	0.0051	0.0072	0.0078	0.0106	0.0200
Intruder shape	0.0125	0.0116	0.0144	0.0151	0.0165	0.0215
Resident + intruder shape	0.0059	0.0123	0.0150	0.0147	0.0154	0.0154
Resident shape	0.0109	0.0128	0.0147	0.0169	0.0173	0.0263

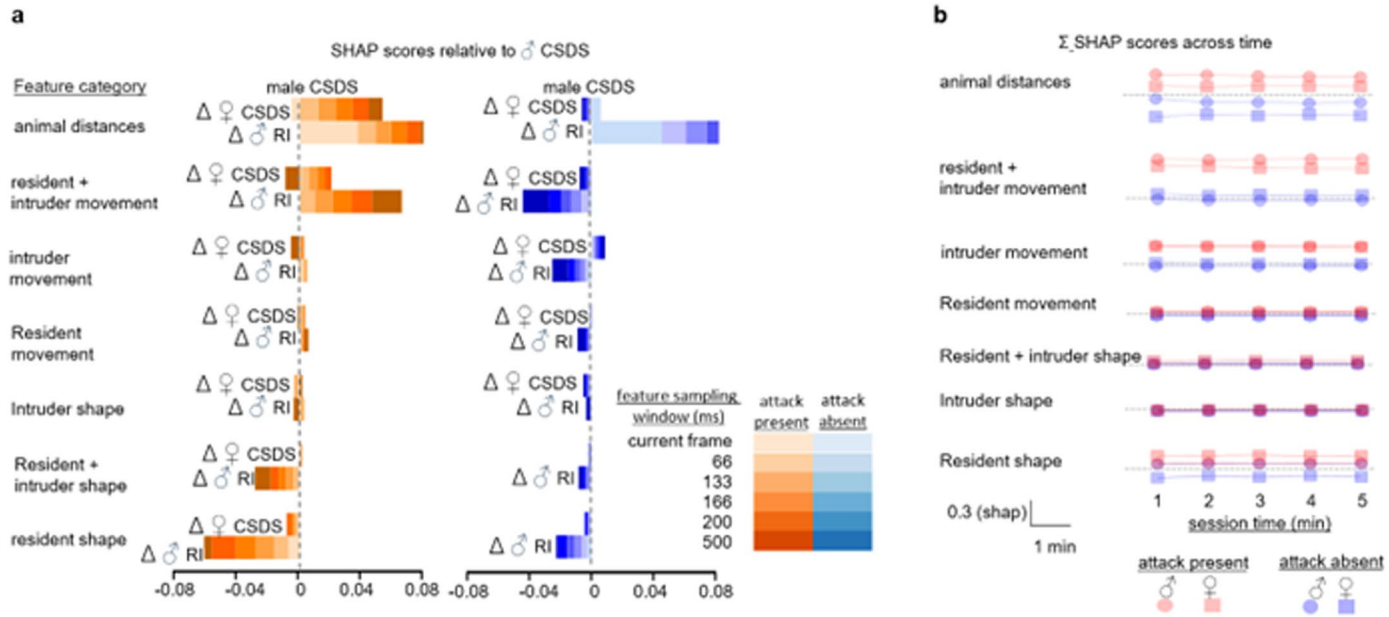
Anogen SHAP (Positive frames)	0ms	66ms	133ms	166ms	200ms	500ms
Animal distances	0.2393	0.0340	0.0336	0.0338	0.0335	0.0292
Intruder movement	0.0023	0.0050	0.0086	0.0106	0.0132	0.0216
Resident + intruder movement	0.0010	0.0054	0.0094	0.0121	0.0169	0.0314
Resident movement	0.0034	0.0053	0.0104	0.0123	0.0151	0.0234
Intruder shape	0.0109	0.0104	0.0111	0.0122	0.0128	0.0184
Resident + intruder shape	0.0037	0.0075	0.0081	0.0081	0.0086	0.0114
Resident shape	0.0201	0.0151	0.0168	0.0172	0.0178	0.0217

Defensive SHAP (Positive frames)	0ms	66ms	133ms	166ms	200ms	500ms
Animal distances	0.1306	0.0618	0.0581	0.0532	0.0461	0.0417
Intruder movement	0.0039	0.0042	0.0052	0.0060	0.0063	0.0174
Resident + intruder movement	0.0043	0.0168	0.0234	0.0253	0.0298	0.0556
Resident movement	0.0045	0.0031	0.0041	0.0047	0.0061	0.0149
Intruder shape	0.0164	0.0133	0.0130	0.0127	0.0125	0.0128
Resident + intruder shape	0.0027	0.0057	0.0067	0.0074	0.0075	0.0081
Resident shape	0.0059	0.0101	0.0135	0.0146	0.0154	0.0203

Escape SHAP (Positive frames)	0ms	66ms	133ms	166ms	200ms	500ms
Animal distances	0.0449	0.0104	0.0121	0.0125	0.0139	0.0253
Intruder movement	0.0237	0.0405	0.0585	0.0615	0.0669	0.0582
Resident + intruder movement	0.0030	0.0158	0.0327	0.0458	0.0526	0.0516
Resident movement	0.0021	0.0021	0.0025	0.0030	0.0037	0.0068
Intruder shape	0.0107	0.0109	0.0131	0.0137	0.0134	0.0210
Resident + intruder shape	0.0038	0.0088	0.0099	0.0097	0.0091	0.0094
Resident shape	0.0036	0.0057	0.0067	0.0071	0.0074	0.0109

Extended Data Fig. 7 | SHAP values for positive frames of UW mouse classifiers used in Figs. 5, 6. SHAP values for attack, pursuit, anogenital sniffing, defensive, and escape behavioral classifiers used in Figs. 5, 6.

https://doi.org/10.1038/s41593-024-01649-9



Extended Data Fig. 8 | Attack SHAP values across groups and throughout testing sessions. We calculated SHAP values for 1250 attack frames and 1250 non-attack frames within each experimental protocol. (**a**) We used these values to calculate delta shap values, where we evaluated the female CSDS and male RI SHAP values against male CSDS SHAP value baseline. The SHAP analyses revealed large similarities in how feature values affected attack classification probabilities in the three experiments (all feature sub-category delta shap < 0.044). The most notable experiment difference was the importance of animal distance features within the current frame, which was associated with higher attack classification probabilities in the RI experiment than in the male CSDS experiment. Attack classification probabilities in the RI experiments were also less affected by features of the resident shape than in the males CSDS experiment. These differences may relate to the different attack strategies and experimental setup used in the experimental protocols. (**b**) Next, we analyzed SHAP vales for classifying attack and non-attack events in the male and female CSDS experiments within 1 min bins and showed that SHAP values are not affected by time of session.

nature portfolio

Corresponding author(s): Sam A. Golden

Last updated by author(s): Mar 8, 2024

Reporting Summary

Nature Portfolio wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Portfolio policies, see our Editorial Policies and the Editorial Policy Checklist.

Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

n/a	Confirmed	
☐	☒	The exact sample size (*n*) for each experimental group/condition, given as a discrete number and unit of measurement
☐	☒	A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
☐	☒	The statistical test(s) used AND whether they are one- or two-sided *Only common tests should be described solely by name; describe more complex techniques in the Methods section.*
☐	☒	A description of all covariates tested
☐	☒	A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
☐	☒	A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)
☐	☒	For null hypothesis testing, the test statistic (e.g. *F*, *t*, *r*) with confidence intervals, effect sizes, degrees of freedom and *P* value noted *Give P values as exact values whenever suitable.*
☒	☐	For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
☒	☐	For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
☐	☒	Estimates of effect sizes (e.g. Cohen's *d*, Pearson's *r*), indicating how they were calculated

Our web collection on statistics for biologists contains articles on many of the points above.

Software and code

Policy information about availability of computer code

Data collection	DeepLabCut (version 1.0) SimBA (versions 1.1.0-1.65.3) Pylon Camera Software (7.4.0) TreeSHAP (0.35.0)
Data analysis	GraphPad Prism 8.0.1, Python 3.6

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors and reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Portfolio guidelines for submitting code & software for further information.

Data

Policy information about availability of data

All manuscripts must include a data availability statement. This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A description of any restrictions on data availability
- For clinical datasets or third party data, please ensure that the statement adheres to our policy

All code and data are available at https://github.com/sgoldenlab/simba. The CRIM13 database is available at https://paperswithcode.com/dataset/crim13.

Research involving human participants, their data, or biological material

Policy information about studies with human participants or human data. See also policy information about sex, gender (identity/presentation), and sexual orientation and race, ethnicity and racism.

Reporting on sex and gender	N/A
Reporting on race, ethnicity, or other socially relevant groupings	N/A
Population characteristics	N/A
Recruitment	N/A
Ethics oversight	N/A

Note that full information on the approval of the study protocol must also be provided in the manuscript.

Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

☒ Life sciences ☐ Behavioural & social sciences ☐ Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see nature.com/documents/nr-reporting-summary-flat.pdf

Life sciences study design

All studies must disclose on these points even when the disclosure is negative.

Sample size	No statistical methods were used to pre-determine sample sizes, but our sample sizes are similar with previous works in the field (Golden 2011, Golden 2013, Newman 2019)
Data exclusions	No data exclusions were performed.
Replication	We did not replicate behavioral cohorts, and all cohorts had appropriate sample sizes.
Randomization	Animals used for chronic social defeat stress were pre-screened for aggression via resident intruder assays and then split into groups of aggressors based on sex. Males were randomly assigned to chronic social defeat stress or resident intruder groups for the environmental comparison.
Blinding	The analyses were not conducted blindly because they were carried out objectively by SimBA. Training sets for SimBA were representative of both males and females, and of animals from each testing condition.

Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

Materials & experimental systems

n/a	Involved in the study
☒	☐ Antibodies
☒	☐ Eukaryotic cell lines
☒	☐ Palaeontology and archaeology
☐	☒ Animals and other organisms
☒	☐ Clinical data
☒	☐ Dual use research of concern
☒	☐ Plants

Methods

n/a	Involved in the study
☒	☐ ChIP-seq
☒	☐ Flow cytometry
☒	☐ MRI-based neuroimaging

Animals and other research organisms

Policy information about studies involving animals; ARRIVE guidelines recommended for reporting animal research, and Sex and Gender in Research

Laboratory animals	Male CD-1 (strain #022, Charles River Labs (CRL)) or female CFW (strain #024, CRL) white coat colored 12-week-old mice were used as residents. Female CFW residents were pair housed with 10-week-old castrated CFW male mice for the duration of the study to elicit aggression60,61, while male CD-1 residents were singly housed. Sex-matched, black coat colored C57BL6/J (C57; strain #000664, Jackson Labs) mice were used as intruders for all aggression assays. C57 females were > 10 weeks old, and males were > 8 weeks old. We chose to use differently-coated social partners to facilitate subject identification, which is improved by using social pairs with different coat colors. We gave all mice free access to standard food chow and water in all experiments. We housed all mice with enrichment (cotton padding) in standard Allentown clear polycarbonate cages covered with stainless-steel wire lids at least one week prior to experiments, and we maintained them on a reverse 12-h light/dark cycle (light off at 0900 am). Rats. Male Long-Evans rats (Simonsen Laboratories, 120-140 days old) were used as residents. Residents were housed in isolation in standard rat cages for the duration of the study. Male Sprague Dawley rats (bred in house, 60-80 days old) were used as intruders. Intruders were pair housed in standard rat cages. We maintained all animals on a reverse 12-h light/dark cycle (light off at 7am) in temperature and humidity controlled rooms.
Wild animals	This study did not involve wild animals.
Reporting on sex	We have collected and analyzed data from male and female mouse dyads in this manuscript.
Field-collected samples	This study did not include field-collected samples.
Ethics oversight	All experiments were performed in accordance with the Guide for the Care and Use of Laboratory Animals under protocols approved by the local Animal Care and Use Committee at each institution (University of Washington, Tufts University, Stanford University, Columbia University, Washington State University, and the Icahn School of Medicine at Mount Sinai).

Note that full information on the approval of the study protocol must also be provided in the manuscript.

Terms and Conditions

Springer Nature journal content, brought to you courtesy of Springer Nature Customer Service Center GmbH (“Springer Nature”).
Springer Nature supports a reasonable amount of sharing of research papers by authors, subscribers and authorised users (“Users”), for small-scale personal, non-commercial use provided that all copyright, trade and service marks and other proprietary notices are maintained. By accessing, sharing, receiving or otherwise using the Springer Nature journal content you agree to these terms of use (“Terms”). For these purposes, Springer Nature considers academic use (by researchers and students) to be non-commercial.
These Terms are supplementary and will apply in addition to any applicable website terms and conditions, a relevant site licence or a personal subscription. These Terms will prevail over any conflict or ambiguity with regards to the relevant terms, a site licence or a personal subscription (to the extent of the conflict or ambiguity only). For Creative Commons-licensed articles, the terms of the Creative Commons license used will apply.
We collect and use personal data to provide access to the Springer Nature journal content. We may also use these personal data internally within ResearchGate and Springer Nature and as agreed share it, in an anonymised way, for purposes of tracking, analysis and reporting. We will not otherwise disclose your personal data outside the ResearchGate or the Springer Nature group of companies unless we have your permission as detailed in the Privacy Policy.
While Users may use the Springer Nature journal content for small scale, personal non-commercial use, it is important to note that Users may not:

1. use such content for the purpose of providing other users with access on a regular or large scale basis or as a means to circumvent access control;
2. use such content where to do so would be considered a criminal or statutory offence in any jurisdiction, or gives rise to civil liability, or is otherwise unlawful;
3. falsely or misleadingly imply or suggest endorsement, approval , sponsorship, or association unless explicitly agreed to by Springer Nature in writing;
4. use bots or other automated methods to access the content or redirect messages
5. override any security feature or exclusionary protocol; or
6. share the content in order to create substitute for Springer Nature products or services or a systematic database of Springer Nature journal content.

In line with the restriction against commercial use, Springer Nature does not permit the creation of a product or service that creates revenue, royalties, rent or income from our content or its inclusion as part of a paid for service or for other commercial gain. Springer Nature journal content cannot be used for inter-library loans and librarians may not upload Springer Nature journal content on a large scale into their, or any other, institutional repository.
These terms of use are reviewed regularly and may be amended at any time. Springer Nature is not obligated to publish any information or content on this website and may remove it or features or functionality at our sole discretion, at any time with or without notice. Springer Nature may revoke this licence to you at any time and remove access to any copies of the Springer Nature journal content which have been saved.
To the fullest extent permitted by law, Springer Nature makes no warranties, representations or guarantees to Users, either express or implied with respect to the Springer nature journal content and all parties disclaim and waive any implied warranties or warranties imposed by law, including merchantability or fitness for any particular purpose.
Please note that these rights do not automatically extend to content, data or other material published by Springer Nature that may be licensed from third parties.
If you would like to use or distribute our Springer Nature journal content to a wider audience or on a regular basis or in any other manner not expressly permitted by these Terms, please contact Springer Nature at

onlineservice@springernature.com