A circadian behavioral analysis suite for real-time classification of daily rhythms in complex behaviors

Logan J. Perry[1], Blanca E. Perez[1], Larissa Rays Wahba[2], KL Nikhil[2], Ashley N. Starnes[1], William C. Lenzen[1], and Jeff R. Jones[1,3,4,*]

Measuring animal behavior over long timescales has been traditionally limited to behaviors that are easily measurable with real-time sensors. More complex behaviors have been measured over time, but these approaches are considerably more challenging due to the intensive manual effort required for scoring behaviors. Recent advances in machine learning have introduced automated behavior analysis methods, but these often overlook long-term behavioral patterns and struggle with classification in varying environmental conditions. To address this, we developed a novel pipeline to automate pre-existing machine learning behavior classifiers and a novel attentive long short-term memory sequencing model for robust, long-term analysis of behavior. Our system automatically classifies nine home cage behaviors with a performance approaching that of expert-level human labelers. Critically, classification occurs continuously, across multiple animals, and in real time. As a proof-of-concept, we used our system to record behaviors from 97 mice over two weeks to test the hypothesis that sex and estrogen influence circadian rhythms in home cage behaviors. We discovered novel sex- and estrogen-dependent differences in circadian properties of several behaviors including digging and nesting rhythms. We present a generalized version of our pipeline and novel sequencing model, the "circadian behavioral analysis suite," (CBAS) as a user-friendly, open-source software package that allows researchers to automatically acquire and analyze behavioral rhythms with a throughput that rivals sensor-based methods, allowing for the temporal and circadian analysis of behaviors that were previously difficult or impossible to observe.

[1]Department of Biology, Texas A&M University, College Station, TX; [2]Department of Biology, Washington University in St. Louis, St. Louis, MO; [3]Institute for Neuroscience, Texas A&M University, College Station, TX; [4]Center for Biological Clocks Research, Texas A&M University, College Station, TX. *email: jjones@bio.tamu.edu

Introduction

Understanding the genetic, neural, and ethological mechanisms that temporally organize behavior is a fundamental goal of fields including circadian biology, neuroscience, and ecology. However, the temporal analysis of behavior has been largely limited to behaviors that can be accurately measured with low latency and at high throughput. For instance, optical and electrical sensors enable such analysis of eating, drinking, and locomotor behaviors. These behaviors have been widely studied (although usually independently) at high temporal resolution for experimental durations of weeks, months, or even years (Schwartz and Zimmerman, 1990; Jud et al., 2005; Pendergast et al., 2013; Yamanaka et al., 2013; Metzger et al., 2020). Other more complex behaviors such as rearing, nesting, and grooming can often be measured simultaneously using video recording and manual behavior scoring by trained human observers (van der Veen et al., 2008; Gaskill et al., 2009; van Oosterhout et al., 2012; Fujita et al., 2017; Robinson-Junker et al., 2018; Shuboni-Mulligan et al., 2021). Over long timescales, however, this method becomes impractical because video acquisition inevitably outpaces human labeling, leading to an ever-increasing latency between data acquisition and data analysis. For example, it may take an observer less than a minute to label behaviors in a minute long video recording but, due to the tedium of the task and the limits of the human attention span, it would take that observer much longer than a week to classify behaviors in a week-long video recording (Segalin et al., 2020; Muller et al., 2021). Consequently, these behaviors have been studied infrequently at low temporal resolution for limited experimental durations, such as hourly over the course of a single day. To better understand how animal behavior changes over time, ethologically relevant behaviors (regardless of "measurability") must be measured in individual animals simultaneously, automatically, and, critically, in real time over multiple days and conditions at high temporal resolution (Peters et al., 2015; Grieco et al., 2021; Kahnau et al., 2023).

Over the last two decades, methods have been developed to classify animal behavior from video recordings, ranging from computer vision algorithms, such as centroid tracking, to machine learning-based approaches that use markerless pose estimation (e.g., DeepLabCut or DLC) or raw pixel values (e.g., DeepEthogram or DEG) to quantify behavior (Mathis et al., 2018; Pereira et al., 2020; Zhang et al., 2020; Bohnslav et al., 2021). While previous studies have used machine learning to analyze temporal variation in more naturalistic "home cage" behaviors, these methods have faced several challenges (Steele et al., 2007; Goulding et al., 2008; Jhuang et al., 2010; Adamah-Biassi et al., 2014; Salem et al., 2015). For instance, existing methods tend to disregard long-range temporal information by simplifying analysis to frame-wise positional and motion values. A more holistic approach is needed to capture the temporal dynamics of both the recorded animal and potentially changing in-scene objects on both short and long time scales (Xie et al., 2017). Additionally, existing methods are often constrained by specific environmental conditions, such as video perspective, lighting condition, or subject coloration, which greatly limits their applicability. The development of an adaptable, condition-agnostic system is therefore essential for robust temporal analysis. Perhaps most importantly, existing methods have not been typically used for real-time behavior analysis and have not been used to analyze behavior on a "circadian" timescale of days to weeks or longer. Together, these challenges have prevented the widespread adoption of machine learning classification to the long-term or circadian analysis of behavior. This is likely exacerbated by the lack of user-friendly tools that facilitate acquisition, training, validation, and analysis of key behavioral metrics. To solve this problem, we developed a streamlined approach that allows users to extend modern deep learning methods to emulate the functionality of traditional sensor-based analyses of behavior.

Here, we introduce a versatile, high-throughput, real-time behavior acquisition and analysis pipeline for the temporal analysis of behavior. To do this, we first trained a recently reported behavior classification model to classify nine home cage behaviors in mice (eating, drinking, rearing, climbing, digging, nesting, resting, grooming, and locomotion (Garner, 2017)) and found that, after rigorous validation, our model's performance did not reach our desired performance thresholds. To improve the classification of the two-stream convolutional features produced by the original model, we developed and trained a novel attentive long short-term memory (A-LSTM) model to attend to complex feature patterns across sequences of feature outputs from the original model (Vaswani et al., 2017; Staudemeyer and Morris, 2019). We found that our A-LSTM model's performance surpassed that of the original model's classifier scheme and approached the performance of expert-level human labelers. Next, we created the software infrastructure to automatically acquire and infer behavior data from video recording streams in real time, in parallel (here, from 24 mice simultaneously) for an essentially unlimited experimental duration (here, for up to two weeks continuously at 10 frames per second). As a proof-of-concept, we used this system and novel model to test the hypothesis that sex and estrogen influence circadian rhythms in home cage behaviors. Previous studies have identified subtle sex differences in wheel-running activity rhythms (Krizo and Mintz, 2014; Joye and Evans, 2022). However, despite the global regulation of behavior by the circadian system, sex differences in other behavioral rhythms have not yet been identified due to technological limitations. Our automatic inference system allowed us to discover novel sex- and estrogen-dependent differences in the phase and amplitude of several behaviors, including, notably, digging and nesting rhythms. Finally, we developed our A-LSTM model and automatic inference software into a user-friendly, open-source Python package called CBAS, the "circadian behavioral analysis suite." CBAS allows researchers to automatically acquire and analyze behavioral rhythms with a throughput that greatly exceeds manual video labeling and rivals sensor-based methods.

Results

Machine learning classification of behaviors approaches human level performance. We first recorded continuous videos of individually-housed mice for >24 h at 10 fps in both a 12 h:12 h light:dark (LD, where dark is defined as dim 850 nm infrared light) cycle and in constant darkness (DD) (**Fig. 1a**). From these videos, we used strict criteria (see Methods) to define and manually label nine ethologically-relevant behaviors that encompass the majority

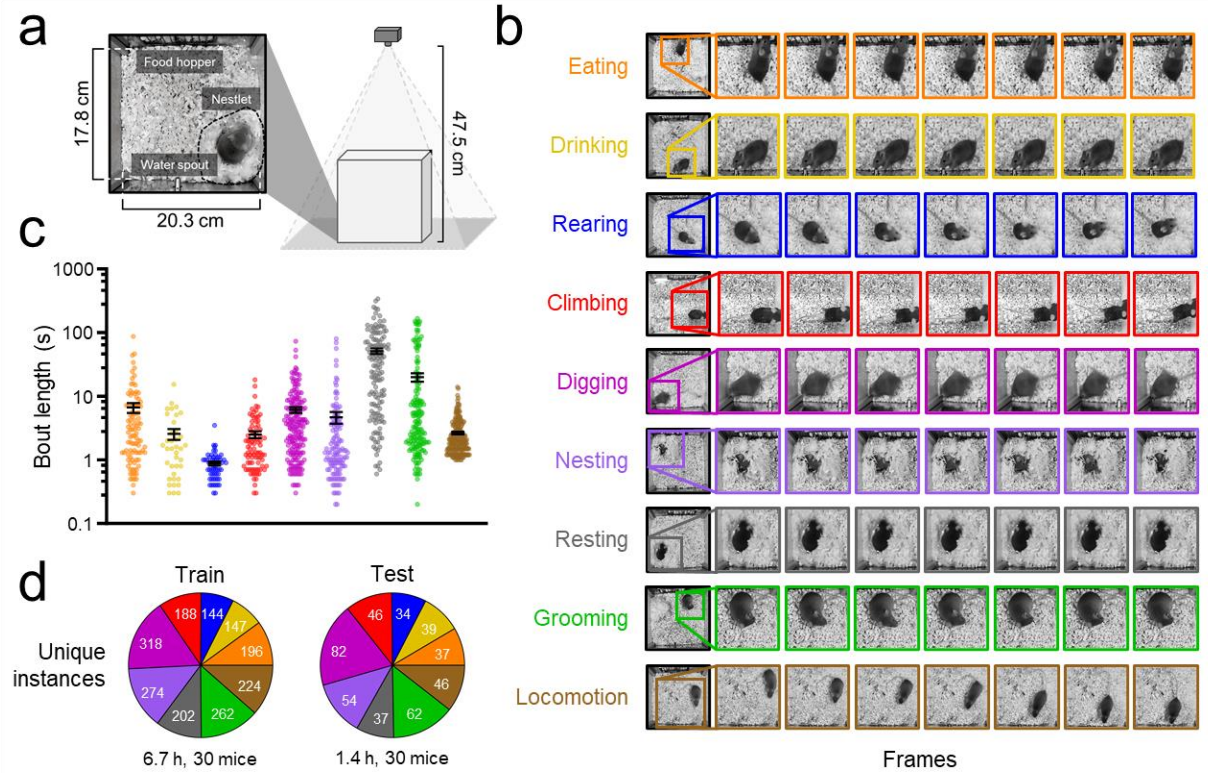

Figure 1. Recording and classification standardization of nine home cage behaviors. a) Schematic of the home cage recording setup. b) Representative examples of individual frames depicting each of nine behaviors (eating, orange; drinking, yellow; rearing, blue; climbing, red; digging, magenta; nesting, purple; resting, gray; grooming, green; locomotion, brown). The first frame depicts the behavior occurring in the full field of view, subsequent frames are zoomed in to better illustrate behaviors. c) Bout length (duration of a behavioral instance) for each behavior within a maximum window size of 360 s. n ≥ 38 bouts from 29 to 30 mice per behavior. Lines and error bars depict mean ± SEM. d) Number of unique instances of each behavior in the 8.1 h human-labeled dataset broken down by training and test sets.

of an individual singly-housed mouse's daily behavioral repertoire, including maintenance, exploratory, and inactive behaviors: eating, drinking, rearing, climbing, digging, nesting, resting, grooming, and locomotion (**Fig. 1b**) (Garner, 2017). For each behavior, we identified the average length of time for a "bout," or behavioral instance (**Fig. 1c**). This allowed us to generate balanced training and test sets from segments of videos sampled from 30 mice that contained a balanced number of unique instances of each behavior (**Fig. 1d**). To control for lighting conditions, we sampled video segments such that there was a balanced representation of each behavior during both the animal's active and inactive phases.

We next used these training and test sets to train a supervised deep learning model, DeepEthogram (DEG) (Bohnslav et al., 2021). DEG identifies behaviors based on a video frame's raw pixel values and produces a binary output matrix – an "ethogram" – that indicates if a behavior is present or absent in a given frame. We reasoned that this temporally sequenced ethogram output would be ideal for quantifying behavioral rhythms because it could be readily analyzed using field standard circadian analysis methods that are optimized for time series data. Briefly, our DEG model comprised three machine learning models: a "flow generator" that estimated optic flow across video frames, a "feature extractor" that determined the probability of a behavior being present on a given frame based on a low-dimensional set of temporal and spatial features, and, finally, a "sequence model" that further refined model predictions using a temporal gaussian mixture (TGM) model with a larger temporal receptive field.

If we wanted to use our model to automatically infer days of video – millions or potentially billions of frames that would never be seen by a human – it was critical that our model was extensively validated. Model performances quantified across all measured behaviors of existing commercial (e.g., HomeCageScan) and non-commercial methods used for the temporal analysis of home cage behaviors are either unreported or, typically, mediocre. Thus, after training our model, we performed rigorous validation of the model's predictions on our labeled test sets with stringent model performance thresholds. Importantly, we did not adjust model hyperparameters based on our model's test set performance. Our predetermined criteria for a "successful" model was a score of at least 0.80 for each behavior frame on multiple demanding metrics: F1 score, normalized Matthews correlation coefficient (nMCC), and balanced accuracy (see Methods) (Brzezinski et al., 2020; Chicco and Jurman, 2020; Grandini et al., 2020). A successful model would also ideally meet or exceed the performance of a trained human classifier labeling the same test set. Surprisingly, although our DEG model performance was excellent on behaviors locked to a spatial region (eating, drinking, climbing) or with minimal motion (resting), its

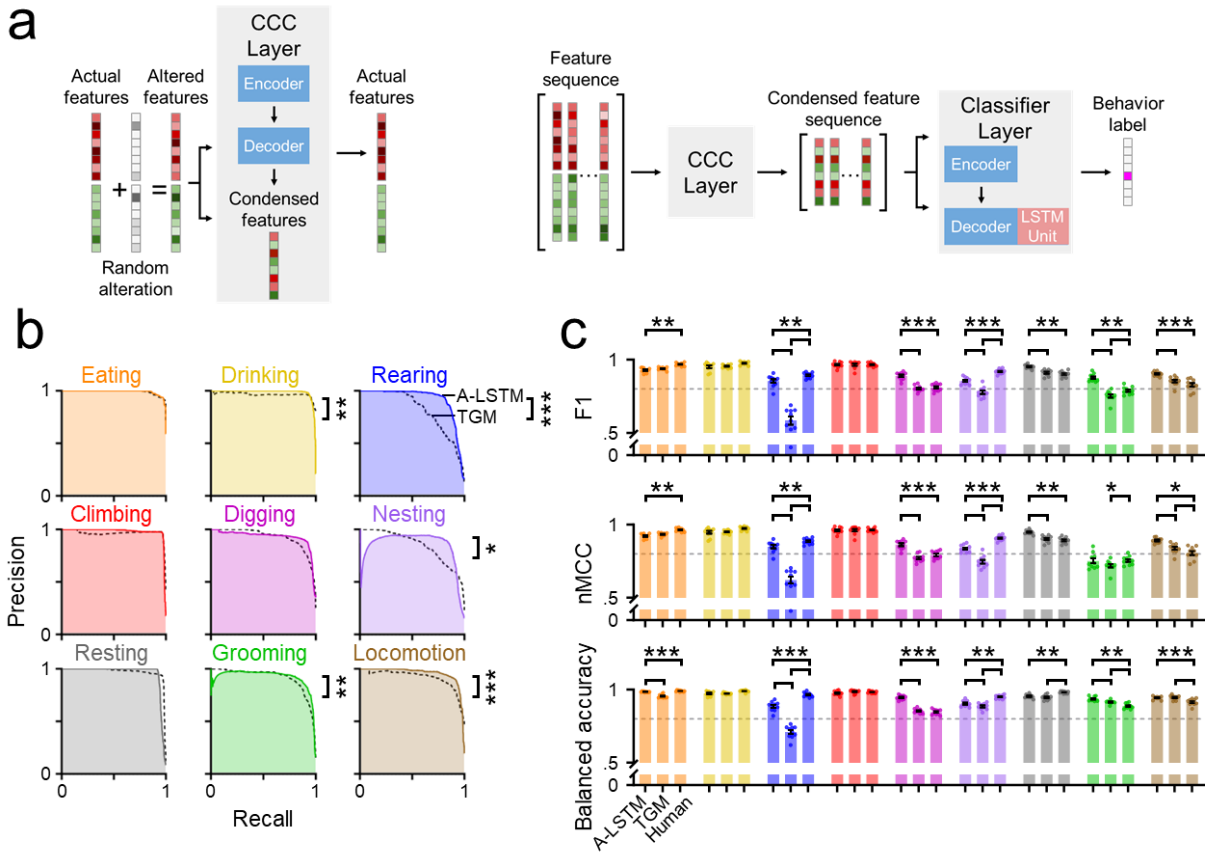

Figure 2. An attentive long short-term memory sequence model approaches expert-level performance on behavior classification. a) Schematic of the attentive long short-term memory (A-LSTM) sequence model. *Left*, autoencoder layer that reconstructs flow (red) and spatial (green) features from randomly-altered input feature vectors. *Right*, classification layer that takes a variable length sequence of condensed feature vectors (red and green) as an input and outputs a behavioral label (magenta) for each sequence. CCC, "clean, condense, combine" autoencoder; LSTM, long short-term memory. b) Precision-recall curves for each behavior calculated for the A-LSTM sequence model (colored lines) and the TGM sequence model (dashed lines) by varying the decision threshold of each binary classifier. Shading depicts the area under the precision-recall curve (AUPRC) for the A-LSTM sequence model, which was compared with the AUPRC for the TGM model. Bootstrap test, *, $p < 0.05$; **, $p < 0.001$; ***, $p < 0.0001$. c) Performance metrics for each behavior calculated for the A-LSTM sequence model, the TGM sequence model, and a trained human classifier. n = 10 sets of 1000 randomly sampled test set frames per behavior. Dashed line depicts a predefined performance threshold of 0.80. Lines and error bars depict mean ± SEM. F1, F1 score; nMCC, normalized Matthews correlation coefficient. Two-way ANOVA with post-hoc Tukey's multiple comparisons test, *, $p < 0.05$; **, $p < 0.01$; ***, $p < 0.001$.

performance was poor on behaviors with more complex motions (rearing, digging, nesting, and grooming, **Figs. 2b,c**).

To address the failure of our original model in identifying complex behaviors, we replaced the TGM sequence model layer of our DEG model with a novel motion sequencing model (attentive long short-term memory or "A-LSTM," **Fig. 2a**) (Vaswani et al., 2017; Staudemeyer and Morris, 2019). Briefly, this alternative sequence layer comprises two attentive encoder/decoder based models. The first of these is an autoencoder that cleans, condenses, and combines (CCC) the raw spatial and flow output features generated from the DEG feature extractor model. We trained this autoencoder to reconstruct true flow and spatial features from randomly altered input feature vectors from ~14,000,000 frames of video data. Next, we constructed a classification layer based on a similar architecture to the CCC layer with an added long short-term memory (LSTM) unit for sequence classification. This layer takes a variable length sequence of condensed feature vectors (here, a maximum length of fifteen vectors or 1.5 seconds of video) and outputs a one-hot encoded behavioral label for each sequence.

We then compared the performance of our A-LSTM model with performance of both our previous TGM sequence model and that of a trained human classifier. Each of these groups (A-LSTM, TGM, human) were given a 15 frame (1.5 s) window to predict behaviors from our test set. Because our performance metrics (F1 score, nMCC, and balanced accuracy) require us to define a specific threshold value at which behavior probabilities are converted into a binary prediction, we also generated precision-recall curves across different probability thresholds for A-LSTM and TGM sequence models. We did not generate human classifier precision-recall curves because in our training set human labels are inherently binary, not probabilistic. A-

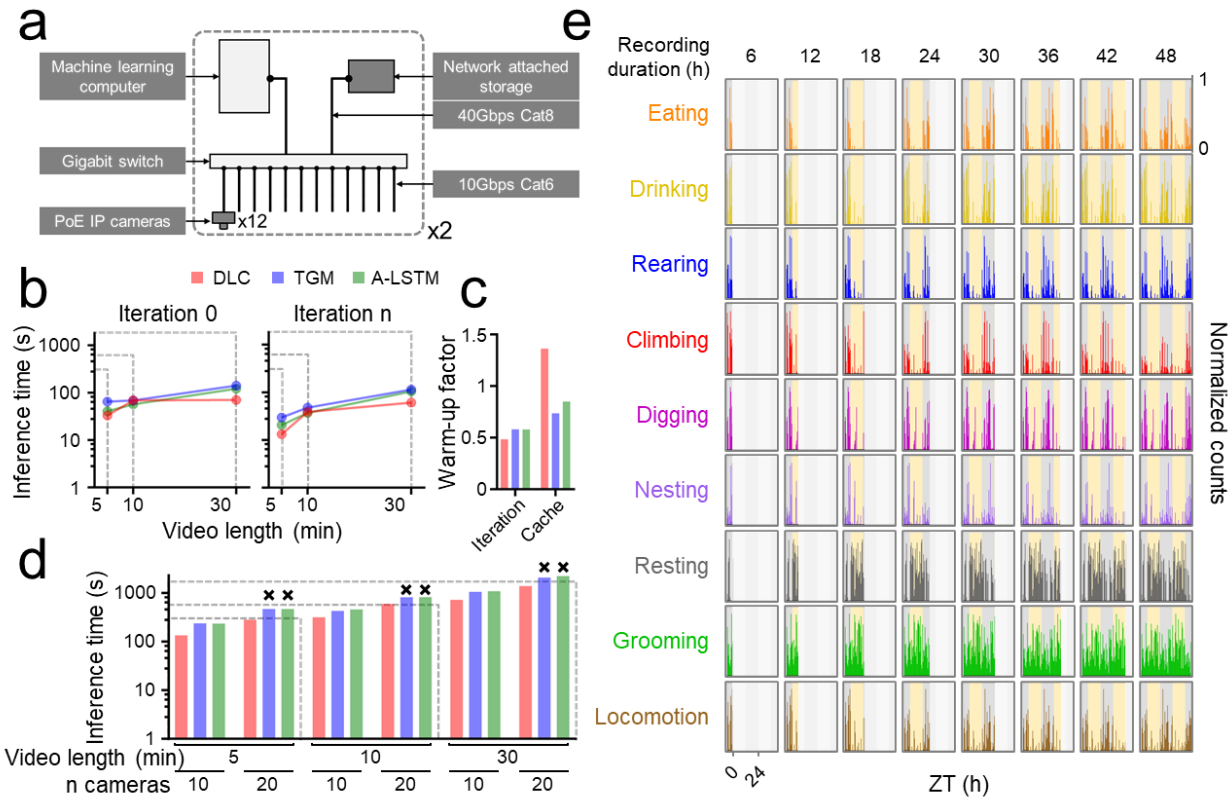

Figure 3. A-LSTM allows for real-time behavior classification. a) Schematic of the real-time video recording, processing, and inferring system comprising two sets of 12 PoE (power over ethernet) IP cameras networked to a switch that passes streaming video data to a machine learning computer for video inference and a network-attached storage device for video backup. b) Single-video inference times for video segments of various lengths calculated for the A-LSTM sequence model (green), the TGM sequence model (blue), and a skeletal pose estimation model without behavior classification (red, DLC). Dashed lines depict the times at which inference time equals the length of the video segment; above this line real-time inference fails. *Left*, inference times for the first inference iteration in a repeated inference test. *Right*, average inference times (mean ± SEM) for the subsequent three inference iterations. c) Warm-up factors calculated for each model averaged across time steps of 5 min, 10 min, and 30 min. Iteration, inference time for the first inference iteration divided by the average inference time for the subsequent three inference iterations. Cache, actual inference time of a video segment divided by the projected time based on the inference time of a shorter video segment. d) Inference times for combinations of video segment length and number of cameras used to simultaneously stream video segments calculated for each model. Dashed lines depict the times at which inference time equals the length of the video segment. Failure of real-time inference for a particular combination of segment length, camera number, and inference model is represented by a black X above the bar. e) Representative activity profiles for each behavior from an individual mouse recorded in a 12 h:12 h light:dark (LD) cycle for 48 h. 30 min segments of continuously recorded video were automatically processed, inferred, and plotted over the duration of the recording, "filling in" over time. For visualization, plots shown here are only updated every 6 h. ZT, zeitgeber time.

LSTM greatly outperformed TGM on most metrics. We found that the areas under the precision-recall curves (AUPRC, a summarization of model performance as a function of probability cut-off threshold) for our A-LSTM model were significantly greater than for our TGM model for drinking, rearing, nesting, grooming, and locomotion behaviors (**Fig. 2b**). A-LSTM also met or exceeded our predefined performance threshold of 0.80 on all performance metrics, whereas our TGM model failed to meet this F1 and nMCC score threshold for rearing, digging, nesting, and grooming behaviors (**Fig. 2c**). Notably, A-LSTM exhibited greater performance than TGM even on behaviors that had F1, nMCC, or balanced accuracy scores of > 0.80, such as resting, locomotion, and eating behaviors. Our A-LSTM model also met or exceeded human classifier performance on metrics for most behaviors, including drinking, climbing, digging, resting, grooming, and locomotion. Together, these results demonstrate that our A-LSTM sequence model's performance on our test set approaches that of expert-level human classifiers. These performance results confidently indicate a high level of reliability of our model, which would allow us to perform behavior inference on a circadian timescale of days to weeks.

Behavior classification occurs in real time. Regardless of our A-LSTM model's exceptional performance, the complexity of machine learning models often translates into poor usage speeds and low throughput in practice. Using A-LSTM as an enhancement to (or replacement for) traditional sensor-based behavior analysis requires us to use it to infer

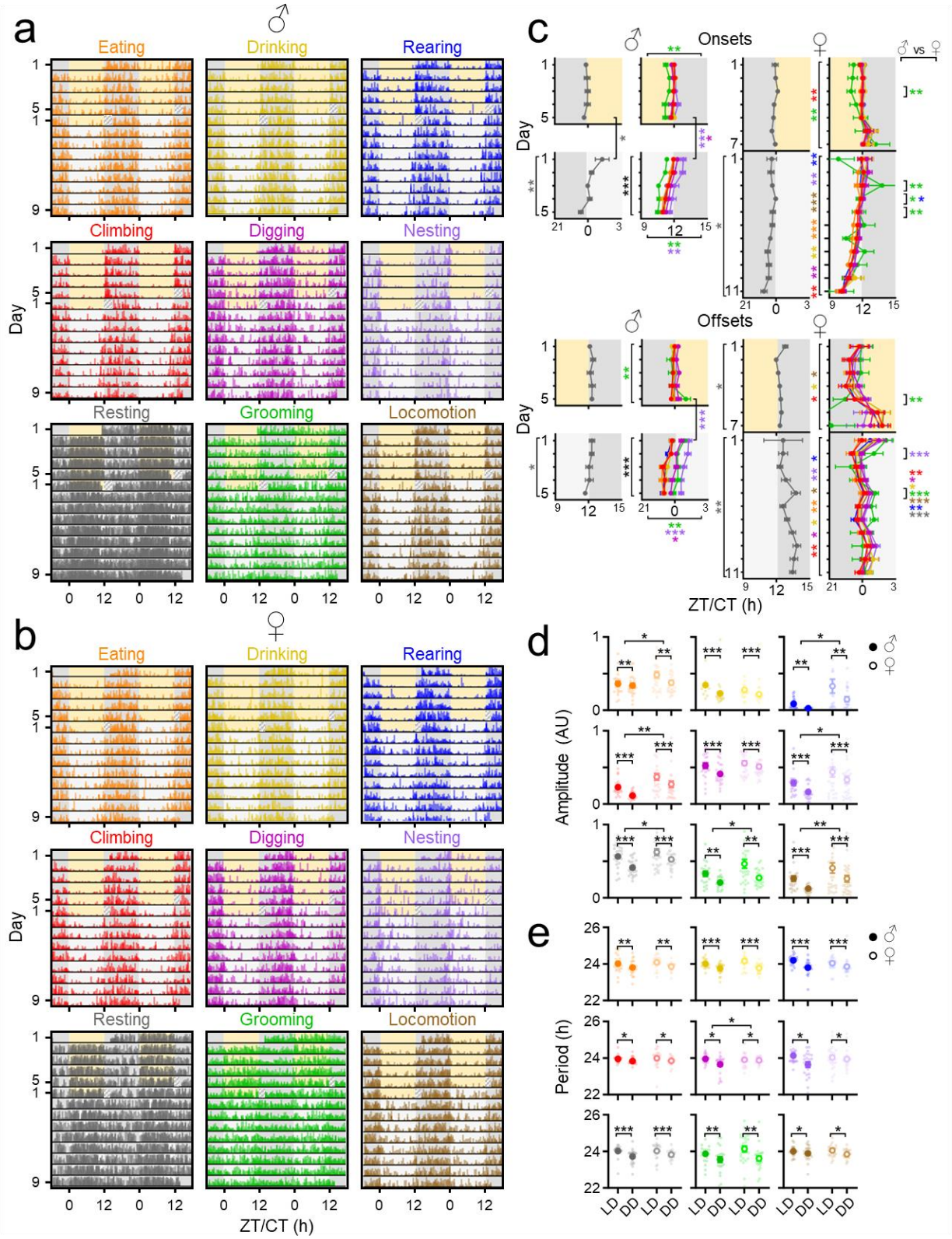

Figure 4. Male and female mice exhibit distinct circadian rhythms in home cage behaviors. a,b) Representative double-plotted actograms depicting behaviors (colored lines on each row) averaged across four male mice or four female mice that started the experiment in the same estrous state recorded over 5 d in a 12h:12h light:dark (LD) cycle (gray and yellow shading) and 9 d in constant darkness (DD; gray and light gray shading). Gray hatched boxes indicate times at which video recording was paused; importantly, this did not affect the ongoing behavior of the mouse, which continued without interruption. c) Phase maps depicting the times of activity

Figure 4 (cont.). onsets (top) and the times of activity offsets (bottom) for each behavior (colored circles and connecting lines) for male (n = 24) and female (n = 27) mice on each day of LD (gray and yellow shading) and DD (gray and light gray shading). Male mice were analyzed over 5 d in LD and 5 d in DD because only a subset of male mice was recorded for a full 9 d in DD. Female mice underwent estrous staging prior to beginning recording, which allowed us to sort them into groups in order to adjust their start times such that plots were aligned by their first day of proestrus. Because of this adjustment, female mice were recorded over 5 d in LD and 9 d in DD but were analyzed over 7 d in LD and 11 d in DD. Brackets on the phase maps indicate significance: brackets on the left depict differences over time; brackets on the top or bottom depict differences between behaviors; brackets linking LD and DD phase maps depict differences across light cycles; and brackets on the far right depict days when behavior phases differ between male and female mice. Asterisk color depicts the behavior(s) that are significantly different: orange, eating; yellow, drinking; blue, rearing; red, climbing; magenta, digging; purple, nesting; gray, resting; green, grooming; brown, locomotion; black, all behaviors. Two-way repeated measures ANOVA (male mice) or mixed effects model (female mice) with post-hoc Tukey's multiple comparisons test; *, $p < 0.05$; **, $p < 0.01$; ***, $p < 0.001$. d,e) Normalized amplitude of a fitted sine wave and period for each behavior rhythm for male (filled circles) and female (open circles) mice measured across all days in LD and across all days in DD. Two-way repeated measures ANOVA with post-hoc Tukey's multiple comparisons test; *, $p < 0.05$; **, $p < 0.01$; ***, $p < 0.001$.

videos in real time. That is, a video clip of n second duration must be recorded, processed, and automatically inferred by A-LSTM before n seconds have elapsed and the next video segment is ready to be processed and inferred. To match the high throughput of sensor-based analysis (e.g., many running wheels can be recorded in parallel), we also need to be able to record, process, and automatically infer behaviors from videos recorded from multiple mice simultaneously. To solve this problem, we developed a hardware and software pipeline that allowed us to automatically and continuously record and infer behaviors in real time from up to 24 mice in parallel (**Fig. 3a**). Our system comprises power over ethernet (PoE) IP cameras connected in parallel to Gigabit switches. These switches stream video data that is binned into constant length time segments onto dedicated machine learning computers for inference and network-attached storage devices for backup.

Before using our system, we needed to identify the video segment length (in minutes) such that video data from x cameras can be inferred within that temporal window. To do this, we first calculated the single camera inference times for several potential models including our TGM sequence model, our A-LSTM sequence model, and, for comparison, a skeletal pose estimation model without behavior classification (DLC) (**Fig. 3b**) (Mathis et al., 2018). We found that all models were able to infer video data from a single camera within each of the temporal windows tested. Because we observed that with repeated inference the first inference iteration took noticeably longer than subsequent iterations, we next calculated the "warm-up factor" for each model averaged across successive video segments (**Fig. 3c**). We defined the warm-up factor as the inference time for the first iteration divided by the average inference times of three subsequent iterations. We found that because of memory caching our TGM and A-LSTM sequence models performed more efficiently as the time segment length increased. For example, the inference time of a 10 min video segment is shorter than the projected inference time based on doubling the inference time of a 5 min video segment. Surprisingly, we did not observe this with the DLC model, most likely because DLC is a frame-by-frame skeletal model that disregards temporal dynamics for pose estimation.

Next, to apply real-time inference to multiple animals in parallel, we first needed to identify the maximum number of cameras that could infer behaviors simultaneously within a reasonable video segment length. We used our TGM sequence model, our A-LSTM sequence model, and our DLC pose estimation model to calculate behavior inference times for various combinations of time segment lengths (5 min, 10 min, or 30 min) and numbers of cameras used to simultaneously stream video segments (10 or 20; **Fig. 3d**). We found that all models were able to infer video data from 10 cameras simultaneously regardless of video segment length. However, when our TGM or A-LSTM sequence models were used to infer video data from 20 cameras simultaneously, inference time exceeded the length of the video segment regardless of video segment length. This indicated a failure of real-time inference. Based on these results (and our experimental setup in which our behavior cabinets can hold a maximum of 12 mouse cages each), we chose to proceed with using our A-LSTM sequence model to infer videos with a video segment length of 30 min on a system comprising two sets of 12 cameras networked to individual machine learning computers (**Fig. 3a**). To test the efficacy of our system, we recorded videos from 24 mice simultaneously over 48 h in a 12h:12h LD cycle (**Fig. 3e**). Our system was able to successfully process video clips, infer behaviors, and plot time series activity profiles for each behavior over the duration of the recordings, "filling in" over time similarly to how wheel-running activity profiles are plotted by commercial circadian activity monitoring software. Together, these results demonstrate that our A-LSTM model can be used to automatically and continuously classify behaviors from multiple animals for an essentially unlimited experimental duration.

Sex influences circadian rhythms in home cage behaviors. Next, we applied our A-LSTM sequence model and automatic inference system to a fundamental question in circadian biology: how do sex and estrogen influence circadian rhythms in behavior? Subtle sex differences in wheel-running activity rhythms have been previously identified (Lee et al., 2004; Kuljis et al., 2013, 2016; Krizo and Mintz, 2014; Joye and Evans, 2022; Anderson et al., 2023). However, because of technological constraints, whether (and how) males and females differ in other behavioral rhythms is unknown. To address this problem, we continuously recorded videos, inferred behaviors, and generated actograms (a field-standard method of plotting activity profiles over multiple days) from male (n = 24) and female (n = 27) mice over 5 d in LD and over 5 to 9 d in DD (**Figs. 4a, b**). Female mice underwent estrous staging prior to beginning recording, allowing us to sort them into groups

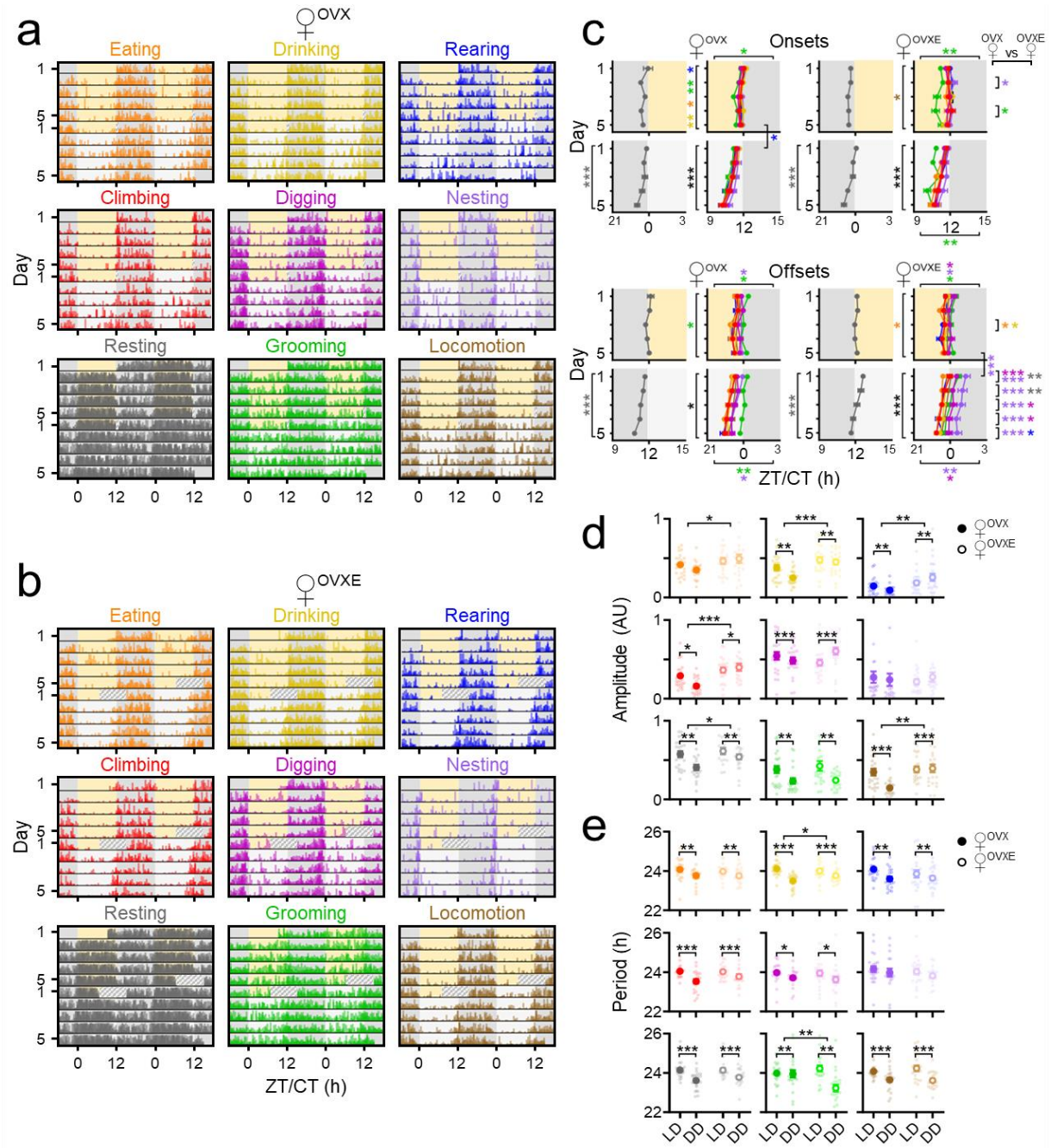

Figure 5. Ovariectomized and ovariectomized, estradiol-supplemented female mice exhibit distinct circadian rhythms in home cage behaviors. a,b) Representative double-plotted actograms depicting behaviors (colored lines on each row) averaged across four ovariectomized (OVX) female mice or four ovariectomized, estradiol-supplemented (OVXE) female mice recorded over 5 d in a 12h:12h light:dark (LD) cycle (gray and yellow shading) and 5 d in constant darkness (DD; gray and light gray shading). Gray hatched boxes indicate times at which video recording was paused; importantly, this did not affect the ongoing behavior of the mouse, which continued without interruption. c) Phase maps depicting the times of activity onsets (top) and the times of activity offsets (bottom) for each behavior (colored circles and connecting lines) for OVX (n = 24) and OVXE (n = 22) mice on each day of LD (gray and yellow shading) and DD (gray and light gray shading). Brackets on the phase maps indicate significance: brackets on the left depict differences over time; brackets on the top or bottom depict differences between behaviors; brackets linking LD and DD phase maps depict differences across light cycles; and brackets on the far right depict days when behavior phases differ between OVX and OVXE mice. Asterisk color depicts the behavior(s) that are significantly different: orange, eating; yellow, drinking; blue, rearing; red, climbing; magenta, digging; purple, nesting; gray, resting; green, grooming; brown, locomotion; black, all behaviors. Two-way repeated measures ANOVA with post-hoc Tukey's multiple comparisons test; *, $p < 0.05$; **, $p < 0.01$; ***, $p < 0.001$. d,e) Normalized amplitude of a fitted sine wave and period for each behavior rhythm for OVX (filled circles) and OVXE (open circles) mice measured across all days in LD and across all days in DD. Two-way repeated measures ANOVA with post-hoc Tukey's multiple comparisons test; *, $p < 0.05$; **, $p < 0.01$; ***, $p < 0.001$.

adjusted such that their actograms were aligned by their first day of proestrus. Because of this estrous state alignment, female mice were recorded over 5 d in LD and 9 d in DD but their actograms were plotted and analyzed over 7 d in LD and 11 d in DD. We used these actograms to quantify key circadian properties of each behavioral rhythm including phase, amplitude, and period (**Figs. 4c-e**).

First, we measured two distinct phase markers for each behavior, the times of activity onset and offset (**Fig. 4c**). In LD, male mice exhibited stable phase angles of entrainment for the onsets and offset of each behavior except grooming. All behaviors except resting and grooming began around the light:dark transition (ZT or zeitgeber time 12, where ZT 0 is defined as lights on) and ended around the dark:light transition, ZT 0. As expected for nocturnal animals, resting began around ZT 0 and ended around ZT 12. Curiously, grooming had a longer alpha (duration of activity) in LD, beginning about 1 h before (ZT 11), but ending at the same time as (ZT 0), other non-resting behaviors. Female mice in LD exhibited stable phase angles of entrainment for the onset of resting behavior. However, the offset of resting, and the onsets and/or offsets of multiple behaviors including grooming, climbing, eating, and locomotion, exhibited a phase delay on the last two days of LD (projected proestrus or estrus). That is, despite being entrained to the LD cycle, these female mice began and ended these behaviors 1 to 2 h later than on projected metestrus or diestrus. The onset and offset times of grooming behavior in female mice in LD were also more unpredictable than in males and differed significantly from males on several days of recording.

For male mice, each behavior began and ended slightly earlier on each day in DD as expected for animals with a free-running period of less than 24 h (**Fig. 4c**). Surprisingly, on the first day of DD, nesting, digging, and resting behaviors began about 1 to 2 h later than they began on the last day of LD. Nesting also ended 1 to 2 h later than it ended on the last day of LD. When analyzed throughout DD, we found that nesting and digging behaviors had distinct phases (a later onset for nesting and later offsets for both nesting and digging) compared to all behaviors except grooming. Male grooming behavior in DD had an even longer alpha than in LD, beginning about 1 h before (CT or circadian time 11) and ending about 1 h after (CT 1) all other non-resting behaviors. For female mice, each behavior also began slightly earlier on each day in DD. Interestingly, all behaviors except grooming exhibited a characteristic offset phase delay on the day of proestrus (or projected proestrus) such that they ended about 1 to 2 h later than on other days in DD. Consequently, during proestrus, the phase of most behaviors differed significantly from that of males. The offset times of nesting did not, however, differ between males and females during proestrus because nesting offsets are normally delayed in males. Together, these results demonstrate that behavior rhythms in male and female mice exhibit distinct phase profiles. Specifically, we found that estrous state fundamentally alters behavior phase in female mice and that, in DD, nesting and digging behaviors are significantly delayed in male mice.

We next measured the amplitudes of each behavior rhythm in males and females by fitting a sine wave to the averaged activity profiles across all days in both LD and DD (**Fig. 4d**). We found that the amplitudes of all behavior rhythms in both male and female mice were dampened in DD compared to LD, consistent with prior reports describing how light cycle influences wheel-running activity amplitude (Li et al., 2006; Pasquali et al., 2010). We also observed that the amplitudes for most behavior rhythms (eating, rearing, climbing, nesting, resting, grooming, and locomotion) were significantly greater in female mice compared to male mice, regardless of light cycle. Finally, we calculated the periods of each behavior rhythm in males and females across all days in both LD and DD (**Fig. 4e**). We found that, as expected for nocturnal laboratory mice, the free-running periods in DD for all behavior rhythms in both male and female mice were shorter than the entrained periods in LD (averaged across all behaviors: males LD 24.02 ± 0.03 h; males DD 23.73 ± 0.04 h; females LD 24.04 ± 0.03 h; females DD 23.82 ± 0.03 h). Surprisingly, sex had little effect on period. Digging in females exhibited a slightly lengthened period in DD, but no other behaviors showed significant period differences. Together, these results demonstrate that biological sex has a profound effect on the amplitude of most behavioral rhythms but has little to no effect on periodicity.

Estrogen replacement phenocopies multiple behavior rhythm changes seen in proestrus female mice. To determine whether these observed sex differences in circadian behavior could be explained by differences in endogenous estrogen levels, we again continuously recorded videos, inferred behaviors, and generated actograms from ovariectomized (OVX; n = 24) and ovariectomized, estradiol-supplemented (OVXE; n = 22) female mice over 5 d in LD and over 5 d in DD (**Figs. 5a, b**) (Ström et al., 2012). To serve as internal controls, OVXE mice were the same individual animals used in the OVX recordings. OVX mice have chronically low levels of estrogen similar to the levels found in male mice or during estrus/metestrus in intact females, and OVXE mice have chronically elevated levels of estrogen similar to the levels found during proestrus in intact females. We used these actograms to again quantify key circadian properties of each rhythm including phase, amplitude, and period (**Figs. 5c-e**).

In LD, OVXE mice exhibited a stable phase angle of entrainment for the onsets of all behaviors except locomotion and the offsets of all behaviors except eating (**Fig. 5c**). However, the onsets of eating, drinking, grooming, and rearing behaviors and the offset of grooming behavior in OVX mice were unstable despite being ostensibly entrained to the LD cycle. Grooming in both OVX and OVXE mice had a longer alpha in LD, beginning about 30 min to 1 h earlier and ending about 15 to 30 min later than all other non-resting behaviors. Similarly, the offset of nesting in OVX mice and the offset of nesting and digging in OVXE mice ended later than all other behaviors. The onsets of nesting and grooming behaviors and the offsets of eating and drinking behaviors also differed significantly between OVX and OVXE mice in LD on several days of recording.

For OVX mice, each behavior began and ended slightly earlier on each day in DD, as expected (**Fig. 5c**). On the first day of DD, rearing began about 15 to 30 min earlier than it began on the last day of LD. When analyzed throughout DD, we found that grooming and nesting behaviors ended later than all other behaviors, about 15 min later for nesting and 1 h later for grooming. For OVXE mice, each behavior also began and ended slightly earlier on each day in DD.

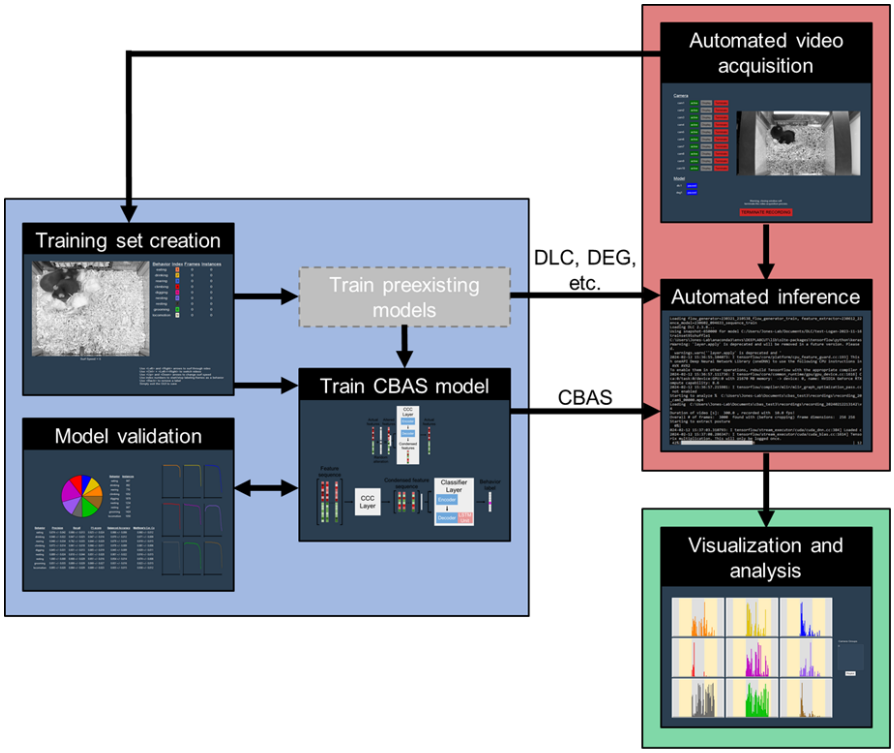

Figure 6. CBAS: a circadian behavioral analysis suite. Schematic of the CBAS pipeline. Blue, training module; red, acquisition module; green, visualization and analysis module.

Surprisingly, on the first day of DD, nesting ended about 1 to 1.5 h later than it ended on the last day of DD. When analyzed throughout DD, we found that grooming behavior started earlier (about 30 min), and digging and nesting behaviors ended later (about 30 min and 1 h, respectively), than other non-resting behaviors. Consequently, on most days in DD, the offset times of nesting and digging were consistently later in OVXE mice compared to the offset times for these behaviors in OVX mice. On several days, other behavior offsets were also delayed in OVXE mice compared to OVX mice, including resting and rearing. Together, these results demonstrate that behavior rhythms in OVX and OVXE mice exhibit distinct phase profiles. Specifically, we found that estrogen replacement significantly extends the duration of several behaviors in DD including, notably, nesting and digging.

We next measured the amplitudes of each behavior rhythm in OVX and OVXE females across all days in both LD and DD (**Fig. 5d**). We found that the amplitudes of all behavior rhythms except eating and nesting in both OVX and OVXE mice were dampened in DD compared to LD. We also observed that the amplitudes for most behavior rhythms (eating, drinking, rearing, climbing, resting, locomotion) were significantly greater in OVXE mice compared to OVX mice, regardless of light cycle. Digging rhythms were also significantly greater in OVXE mice compared to OVX mice in DD, but not in LD. Finally, we calculated the periods of each behavior rhythm in OVX and OVXE mice across all days in both LD and DD (**Fig. 5e**). We found that, as expected, the free-running periods in DD for all behavior rhythms but nesting in both OVX and OVXE mice were shorter than the entrained periods in LD (averaged across all behaviors: OVX LD 24.07 ± 0.02 h; OVX DD 23.70 ± 0.06 h; OVXE LD 24.04 ± 0.04 h; OVXE DD 23.66 ± 0.06 h). Estrogen replacement had little effect on period: only drinking and grooming behaviors in OVXE mice had slightly different periods (longer and shorter, respectively) than in OVX mice. Together, these results demonstrate that estrogen replacement greatly influences both the phase and amplitude of multiple behavior rhythms. Intriguingly, estrogen replacement increases behavior rhythm amplitudes and mimics the delayed behavior offsets we observed in intact female mice during proestrus.

A generalizable circadian behavioral analysis suite. Our A-LSTM model and automatic inference system allowed us to thoroughly investigate how circadian behaviors are influenced by sex and estrogen levels at an unprecedented throughput and acquisition rate. We realized that introducing our software infrastructure to the broader scientific community could be revolutionary to fields seeking to understand the temporal characteristics of animal behavior, including, particularly, circadian biology. We therefore developed a "circadian behavioral analysis suite" (CBAS), a Python package aimed at generalizing the software and A-LSTM model used in our experiments (**Fig. 6**). CBAS is equipped to handle automated, continuous video acquisition, automated inference using existing DEG (with either TGM or A-LSTM sequence models) and DLC models, visualization of behavior actograms in real time, and comprehensive offline circadian analysis. Briefly, CBAS is divided into three modules: an acquisition module, an optional training module, and a visualization and analysis module. The acquisition module is capable of batch processing and performing inference on streaming video data from any number of network-configured real-time

streaming protocol (RTSP) IP cameras. Importantly, CBAS's acquisition module is essentially machine learning model agnostic, allowing for future models to be easily incorporated into the CBAS pipeline. The training module allows the user to create balanced training sets of behaviors of interest, train A-LSTM sequence models for behavior classification, and validate model performance on a naive test set of behavior instances. Finally, the visualization and analysis module displays acquired behavior time series data in real time as actograms that can be readily exported for offline analysis in a file format compatible with ClockLab Analysis, a widely used circadian analysis software. The module also allows for built-in circadian analysis including automatic phase marker identification, calculation of period and amplitude, and behavior transition analysis. Together, these modules present an intuitive, accessible software interface that will allow for the rapid adoption of CBAS by end users with any level of programming ability.

Discussion

Here, we developed and validated a novel system for transforming existing machine learning classifiers into real-time sensors capable of phenotyping circadian rhythms in complex behaviors for an indefinite length of time. We used this pipeline to thoroughly characterize the effects of biological sex and estrogen levels on circadian behavior across 97 individual mouse recordings with a minimum duration of 10 d per recording at 10 frames per second. We then developed this toolkit into an open-source, user-friendly Python package – CBAS – for use by the broader circadian biology community and beyond. CBAS has the potential to reveal temporal variations in behavior that have previously gone undetected in a diverse range of animal models. In addition, CBAS provides scientists with the tools needed to build, adequately validate, and automate highly reliable machine learning classifiers for any complex behavior(s) of interest.

CBAS's extensive model validation, classification power, and customizable, open-source nature set it apart from previous commercial (e.g., HomeCageScan, (Adamah-Biassi et al., 2013, 2014)) and non-commercial (e.g., (Steele et al., 2007; Goulding et al., 2008; Jhuang et al., 2010; Salem et al., 2015)) home cage behavior acquisition tools. For instance, (Jhuang et al., 2010) uses background masking and motion features to classify behaviors with a Hidden Markov Model Support Vector Machine, an outdated, but theoretically capable, architecture. While the authors point to a high classification accuracy, there is little to no information provided regarding more standard machine learning classification metrics such as precision, recall, F1 score, etc. With some assumptions, some of these values can be calculated from the provided confusion matrices, but for most behaviors they fail to meet our strict model performance threshold. Critically, the class balance of the training and test sets used for their model verification is unreported, and neither set remains available to the public. Similarly, (Goulding et al., 2008) fails to perform model performance validation for their supervised learning technique, although they do show that their system is capable of recording behaviors on a circadian timescale. However, because their model only identifies active and inactive behavior states, it is incapable of automatically identifying complex behaviors or animal-environment interactions. Adequate performance metrics, training sets, and testing sets of the HomeCageScan system are likewise scarce, a problem that is exacerbated by its closed-source nature. Importantly, none of these systems has user-friendly customizability to extend classification to other animal models, environments, and novel behaviors. Our goal with CBAS is to allow users with any level of programming ability to integrate completely customizable machine learning models, extensively validate model performance, and record behaviors in real time, indefinitely.

Although there have been several recent advances in supervised and even unsupervised pose-based behavior classification (e.g., SimBA, B-SOiD, A-SOiD, (Nilsson et al., 2020; Hsu and Yttri, 2021; Tillmann et al., 2024)), we prioritized the implementation of DeepEthogram (DEG) because of its capacity to learn specific high-dimensional features from spatial and temporal dynamics derived from raw pixel values (Bohnslav et al., 2021). Admittedly, some supervised pose-based behavior classifiers have demonstrated better performance than "baseline" DEG models at identifying certain behaviors (Tillmann et al., 2024). However, these classifiers lose critical learnable information due to their sweeping dimensional reduction to pure pose dynamics. For example, the subtle home cage environment differences that characterize digging versus nesting behavior in our recording setup would be completely lost in a reduction to pose time series. Instead, we proposed that introducing a sequence model capable of selective multi-head attention to the baseline DEG feature extractor outputs would allow us to access all possible learnable features of a video recording while providing a degree of dimensional reduction that would improve classification performance. Indeed, we found that an attentive long short-term memory (A-LSTM) sequence model achieved a greater utilization of DEG feature extractor outputs through extensive feature condensation and multi-headed attentiveness. This resulted in a dramatic increase in classification power of the DEG feature extractor outputs compared to the originally reported temporal gaussian mixture (TGM) sequence model. We specifically found that our A-LSTM model significantly outperformed the TGM sequence model at classifying behaviors that are characterized by complex movements and behaviors that lack defined spatial orientation features. Importantly, these substantial performance improvements did not significantly affect the behavior classification rate, which allowed us to confidently record all of our behaviors of interest in real time.

In addition to using our novel A-LSTM model for inference, CBAS also leverages our standardized video recording pipeline to allow for real-time, behavior classification that can rival the throughput of sensor-based recording systems (Siepka and Takahashi, 2005; Verwey et al., 2013). CBAS is also incredibly cost effective. The total cost of the hardware we used in this study, including our custom-built mouse cages and circadian behavior cabinets, is only two-thirds the estimated cost of standard commercial systems. Moreover, our setup can be used to record and infer any number of behaviors in parallel, whereas standard circadian acquisition hardware is only capable of recording locomotor behavior by measuring wheel-running activity or infrared beam breaks. Another major advantage of CBAS is its accessibility, presenting a low barrier to entry for the broader circadian research community, including those with

limited programming skills. Additionally, CBAS mirrors the functionality of gold-standard circadian analysis systems by plotting behavior data as actograms in real time, providing immediate feedback about the state of ongoing experiments. CBAS also outputs these behavior actograms in formats compatible with Actimetrics' ClockLab Analysis software, which ensures that researchers can adapt familiar analyses to CBAS-generated behavior data. Notably, CBAS also provides open-source circadian rhythms analysis tools to eliminate the dependency on commercial software to measure fundamental circadian parameters such as phase, period, and amplitude. The open-source and cost effective nature of CBAS essentially democratizes the circadian analysis of complex behaviors, allowing a greater number of researchers to investigate long-term behavior dynamics.

Previous studies have identified numerous sex differences in the temporal patterning of physiology. For example, male and female mice exhibit distinct circadian rhythms in glucocorticoid production, cardiovascular function, body temperature, and immune function (Griffin and Whitacre, 1991; Atkinson and Waddell, 1997; Sanchez-Alavez et al., 2011; Barsha et al., 2016; Walton et al., 2022). However, the question of whether there are pronounced sex differences in the temporal patterning of behavior has, to date, been mostly unanswered. Subtle sex differences have been observed in wheel-running activity rhythms (Lee et al., 2004; Kuljis et al., 2016; Anderson et al., 2023). For instance, male mice show a greater precision of wheel-running activity onsets in LD and female mice show a longer wheel-running activity duration on the day of proestrus in DD (Albers et al., 1981; Kuljis et al., 2013). However, given the presumed global regulation by the circadian system of multiple brain circuits that control distinct behaviors, more work is needed to reveal and understand sex differences in other behavioral rhythms (Starnes and Jones, 2023). We used CBAS to address this by testing the hypothesis that circadian rhythms in behavior differ between males and females. We identified differences in behavioral rhythms including, notably, that nesting and digging rhythms exhibit a distinct phase delay in male, but not female, mice during constant darkness that had not been previously reported. We also observed that most behavioral rhythms in female mice had higher peak-to-trough amplitudes, which is suggestive of more robust circadian organization. This is consistent with previous work showing that the amplitude of wheel-running activity rhythms is greater in female mice compared to male mice (Anderson et al., 2023). Previous studies have also identified that the duration of wheel-running activity is extended (that is, it ends later) on the day of proestrus (Albers et al., 1981). We confirm this finding and extend it to demonstrate that the temporal organization of nearly all behaviors changes across the estrous cycle. These findings reveal critical unseen sex differences in behavior and emphasize the importance of measuring circadian rhythms in behaviors other than locomotor activity.

The limited number of previously identified sex differences in circadian behavior have been speculated to be due to differences in levels of circulating sex hormones and/or sex hormone receptor expression (Walton et al., 2022). Our experiments were designed to allow us to distinguish between differences in behavioral rhythms that are due to biological sex and those that are due to the presence or absence of estrogen. We found that estrogen replacement recapitulates most of our observed sex differences in behavioral rhythms. For example, we found that intact females exhibit higher amplitude circadian rhythms in most behaviors compared to males. Similarly, most behavioral rhythms in ovariectomized, estradiol-supplemented (OVXE) females are more robust than in ovariectomized (OVX) females. Despite the similarities in behavior rhythm amplitudes between intact and OVXE females, OVXE females (with chronically high levels of estrogen) exhibit offsets of nesting and digging behavior rhythms that are more similar to that of males with low levels of estrogen. Why might estradiol supplementation restore behavior rhythm amplitudes to that seen in intact female mice but lengthen the duration of nesting and digging behaviors in DD to that seen in male mice? One possible explanation is the relative distribution of estrogen receptors in brain circuits that regulate different behaviors. Exogenous estrogen has been shown to increase the amplitude of wheel-running activity rhythms through the activation of estrogen receptor (ER)α but to delay the phase of wheel-running activity rhythms through the activation of ERβ (Royston et al., 2014). Indeed, a recent study determined that the lateral hypothalamus (LH), which has subpopulations of both ERα-positive and ERβ-positive neurons, regulates nest-building behavior (Merchenthaler et al., 2004; Sotelo et al., 2022). If these ERβ-expressing LH neurons are preferentially activated during nesting behavior, this could explain why estradiol delays nesting rhythms in ovariectomized female mice. However, this does not explain why male mice (which have low levels of endogenous estrogen) exhibit delayed nesting rhythms, but intact female mice (which have cyclical, low-to-high levels of endogenous estrogen) do not. Further studies will need to determine whether this finding is due to sex differences in developmental circuit wiring, differences in estrogen receptor distribution and/or expression levels, or other factors.

In this study, we used our "circadian behavioral analysis suite" (CBAS) to automatically quantify differences in the circadian regulation of behavior between male and female, and OVX and OVXE female, mice. This approach can be readily expanded to address other critical questions in circadian biology, neuroscience, and ecology, including the ethological investigation of other behavioral rhythms in videos of mice recorded in the laboratory and, potentially, in the wild. Notably, CBAS can also be used for the rapid circadian phenotyping of mice with different genotypes or disorders (Richardson, 2015). Current approaches almost universally measure changes to wheel-running activity rhythms as evidence that a mutation, gene, or drug influences circadian behavior. Here, we found that some, but not all, behavioral rhythms differ by biological sex and by estrogen levels. It is therefore highly likely that any given experimental treatment could cause circadian alterations in behaviors other than, or in addition to, wheel-running activity. CBAS aims to extend the modern toolkit of machine learning classification into any and all long-term behavior assays, greatly expanding the scope of potential hypotheses and impact of future studies.

Methods

Animals. Prior to recording, we group-housed male (n = 24) and female (n = 51; 24 of which were subsequently ovariectomized, see next section) wild-type mice in their home cages in a 12h:12h light:dark cycle (LD,

where lights on is defined as zeitgeber time (ZT) 0; light intensity ~2 x 10^{14} photons/cm^2/s) at constant temperature (~23°C) and humidity (~40%) with food and water provided ad libitum. All mice were between 6 and 12 weeks old at the time of the recording. To determine the estrous stage of female mice, we performed vaginal lavage for four consecutive days prior to beginning long-term recording (Byers et al., 2012). All experiments were approved by and performed in accordance with the guidelines of Texas A&M University's Institutional Animal Care and Use Committee.

Ovariectomy and estradiol capsule implantation. We ovariectomized a cohort of female mice (OVX, n = 24) using standard methods (Ström et al., 2012). Briefly, we made a sterile ~2 cm bilateral incision through the skin and peritoneum immediately dorsal to the ovaries. After ligating and removing each ovary, we sutured the peritoneum and skin incisions. We provided the mice with buprenorphine-SR (1 mg/kg; subcutaneous) and enrofloxacin (0.25 mg/ml; ad libitum in their drinking water) and allowed them to recover in their home cages for at least 1 week prior to beginning long-term recording. To use the OVX mice as internal controls, after recording, we implanted them with a sterile 2 cm silastic capsule containing 17-β-estradiol (36 µg/ml in sesame oil) subcutaneously between the shoulder blades (OVXE; n = 22) (Ström et al., 2012). We excluded two OVX mice from capsule implantation because they had excessive barbering around their ovariectomy incision site. We allowed OVXE mice to recover in their recording cages for 1 d prior to beginning long-term recording.

Experimental housing. We transferred individual mice from their home cages to custom-built recording cages inside custom-built light-tight, temperature-and humidity controlled circadian cabinets for the duration of our experiments. We built the cages (external dimensions, length x width x height: 22.9 cm x 20.3 cm x 21.6 cm; internal dimensions: 20.3 cm x 17.8 cm x 20.3 cm) out of transparent and opaque acrylic panels (thickness, walls and floor: 6.4 mm; lid, 3.2 mm) and T-slot aluminum extrusions (25.4 mm^2) (**Fig. 1a**). We 3D printed custom water bottle holders and food hoppers out of PLA filament, coated them in food safe clear-cast epoxy resin (Alumilite), and affixed them to the acrylic walls. To continue our recordings throughout the dark phase, and to prevent potential glare and shadows from ceiling-mounted lights, we affixed dim infrared (850 nm) light strips to the cage lid using a custom 3D printed cage topper. Prior to recording, we added ~7 mm wood chip bedding and a 25 mm by 50 mm square of cotton nestlet to the cage bottom, added food pellets to the food hopper, and attached a standard water bottle filled with water to the water bottle holder such that its metal spout protruded about 1.5 cm into the cage. Our circadian cabinets were built to hold twelve of our custom-built mouse cages across three vertical shelves, with four cages per shelf. We controlled the ceiling-mounted lights in the cabinets (broad-spectrum white light, ~6 x 10^{13} photons/cm^2/s measured at the cage floor) using ClockLab Data Collection hardware and software (Actimetrics) that communicated via a 5V transistor-transistor logic signal with a high-power power relay (Digital Loggers). We performed daily animal welfare checks using dim red light (650 nm).

Automated video recording. We positioned power-over-internet (PoE) IP cameras without infrared filters (I706-POE, Revotech) equipped with 6 mm lenses (Xenocam) 47.5 cm above the recording cages such that all four corners of the cage, the food hopper, and the water spout were each visible in the recorded video and the mouse and nesting material were in focus. We recorded all videos at 10 frames per second (fps) with in-camera image settings set to a contrast of 130/255, brightness of 140/255, saturation of 0/255, and sharpness of 128/255. We streamed videos at a main stream bitrate of 2048 kilobits per second (kb/s) and a secondary stream bitrate of 256 kb/s. We disabled audio streams to reduce bandwidth. We recorded our mice in cohorts of 8 to 12 mice split between two circadian cabinets. We paused our recordings briefly between light settings (LD and DD) to allow for cage changes, if necessary.

We used FFmpeg, a standard open-source video processing tool, to develop a custom video acquisition system capable of streaming live video and storing successively binned segments of video for each network camera simultaneously. During a recording, FFmpeg automatically handles cropping the video to a region of interest and scaling the video to a desired size (here to a scale of 256x256 pixels). To do this, we connected our cameras in parallel via 10 gigabits per second (Gbps) Cat6 ethernet cables to a Gigabit PoE switch (Aruba JL684A#ABB). We then connected our switches via 40 Gbps Cat8 ethernet cables to our custom machine-learning computers (12-core AMD Ryzen 9 5900X CPU, 32 GB RAM, NVIDIA GeForce RTX 3090 with 24 GB VRAM) (**Fig. 3a**). During a recording, our software creates two threads to monitor the creation of streamed video segments and orchestrate the inference of incoming data (the "storage" and "inference" threads). The storage thread records information about each video segment (creation time, segment length, camera-specific settings), moves the video segments to the corresponding camera directories on the computers, and notifies the inference thread that new videos are available for inference (see below).

Behavior definitions. We defined a list of nine home cage behaviors (eating, drinking, rearing, climbing, digging, nesting, resting, grooming, and locomotion, (Garner, 2017)) with the goal of identifying the key characteristics of each behavior that our two-stream convolutional feature extractor (see below) would be capable of learning (**Supplementary Table 1**). As such, our definitions do not aim to ascribe intent to a behavior (as humans are often inclined to do), but rather contain references to particular features that strictly define behavioral classes. These include spatial features that are necessary constraints on a behavior and temporal features that are split into two groups indicative of the start and stop of a behavior sequence. To further enforce these rigorous criteria defining behaviors, our entire set of training instances were generated by a single labeler.

Model training and inference. To train our baseline DeepEthogram (DEG) classifier, we needed to individually train three components, a "flow generator" that estimates optic flow across video frames, a "feature extractor" that determines the probability of a behavior being present on a given frame based on a low-dimensional set of temporal and spatial features, and a "sequence model" that further refines model predictions using a temporal gaussian mixture (TGM) model with a larger temporal receptive field. We trained our flow generator on a set of videos consisting of approximately 500,000 frames of videos from 8 mice recorded at 10 fps. We then trained our feature extractor using the medium model size preset (deg_m, (Bohnslav et al., 2021)) and our TGM sequence model using a temporal window of 15 frames. We trained both the feature extractor and TGM sequence models on an identical balanced training set used for subsequent training of our attentive long short-term memory (A-LSTM) sequence model (see below). Importantly, we include our model configuration files for all DEG models, our A-LSTM sequence model, and all trained model weights at the Jones lab Github page (see Data Availability section below).

Our A-LSTM sequence model architecture was designed to take as input sequenced outputs from the DEG feature extractor and produce a more robust, frame-to-frame stable, and accurate classification time series (**Fig. 2a**). Our DEG feature extractor model outputs two 512 length vectors for a given video frame: a spatial vector that encodes class-relevant information about the image scene, and a flow vector that encodes class-relevant information about how the image scene changes over an 11 frame window centered on the frame of interest (Bohnslav et al., 2021). To extract pertinent features from these outputs that were not captured by the TGM sequence model, we designed an encoder/decoder unit to selectively attend to patterns within 1024 length combined spatial and flow feature vectors. Our model architecture was heavily inspired by the sequence-to-sequence performance of transformers and long short-term memory models (Hochreiter and Schmidhuber, 1997; Vaswani et al., 2017). During each forward pass of our attentive unit, we linearly transform inputs into the model dimension, add positional encoding, and then apply successive multi-head attention encoding layers to obtain a latent space representation of the input. We take this representation, positionally encode it, and linearly transform it into the model dimension. We then transform this in tandem with the linearly transformed inputs through successive multi-head attention decoding layers to produce logits that are, finally, linearly transformed into an output. These transformations effectively produce a residual attention mechanism capable of encoding critical feature complexities into a condensed form.

In the first "clean, condense, combine" (CCC) layer of our sequence model, we trained the attentive encoder/decoder unit to reconstruct feature vectors from randomly altered versions of the same feature vectors using a stochastic gradient descent method (Adam optimizer) for learning and mean squared error (MSE) as the loss

function, defined as:

$$L_{MSE} = \frac{1}{n}\sum_{j=1}^{n}(y_{oj} - y_{ij})^2$$

where n is the vector length (here, 1024), y_{oj} is the output vector at position j, and y_{ij} is the input vector at position j (Ciampiconi et al., 2023). We specifically designed this task to encourage the model to learn a richer, mixed, and compressed version of the input features. For this task, we identified optimal hyperparameters that minimized the mean squared error (MSE) loss to be a latent dimension of 64, a model dimension of 128, 15 transformer encoder layers, 15 transformer decoder layers, and 64 self/multi-attention heads per transformer encoder/decoder layer. We trained this layer with a constant learning rate of 0.00001 across 5,000 epochs on a training set of approximately 14,000,000 DEG feature vectors. We froze the parameters in this pre-trained autoencoder when we subsequently trained the classification layer.

We designed the classification layer of our sequence model to use a combined sequence vector of length 137 containing 128 logits from the CCC layer and 9 logits (here, a sequence of maximum length 15) from the DEG feature extractor as input. We passed each individual logit vector of this sequence through another attentive encoder/decoder unit to generate a sequence of features that are temporally compressed by a long short-term memory (LSTM) unit. We linearly transformed the resulting logits from the LSTM unit, achieving an output dimension size equal to the number of classes, and, in order to output a prediction probability for each class, passed them through a softmax function defined as:

$$s(y_{oj}) = \frac{e^{y_{oj}}}{\sum_{k=1}^{n} e^{y_{ok}}}$$

where n is vector length (here, 9), y_{oj} is the output vector at position j, and y_{ij} is the input vector at position j.

We next trained this layer on a balanced set of behavior instances sampled across the light and dark phases from 30 unique mice and cages. To provide robustness for classifications at the edges of videos, we randomly shortened one percent of the input sequences and padded them with zero vectors. For this task, we trained our model using a cross entropy loss function, defined as:

$$L_{CE} = -\sum_{j=1}^{n} y_{ij} \log\left(s(y_{oj})\right)$$

where, again, n is vector length (here, 9), y_{oj} is the output vector at position j, and y_{ij} is the input vector at position j. We identified optimal hyperparameters that minimized the cross entropy loss to be a latent dimension of 64, a model dimension of 128, 5 transformer encoder layers, 5 transformer decoder layers, and 16 self/multi-attention heads per transformer encoder/decoder layer. For the LSTM unit, we identified the optimal hyperparameters to be 2 layers and a hidden state size of 128. We trained this layer with a constant learning rate of 0.00001 for 50 epochs using an Adam optimizer.

To discourage model overfitting, we employed a dropout rate of 0.2 after each linear layer of our model, which randomly ablated network outputs to encourage even learning across the network parameters (Ying, 2019). To ensure better generalized learning, we also chose our model hyperparameters following a common machine learning heuristic such that our trainable model parameter size was approximately an order of magnitude smaller than our number of training inputs. Importantly, having a higher number of model parameters increases model complexity but can lead to "overfitting," or rote memorization, if the training set is not large or diverse enough (Lever et al., 2016). Because our input vectors and pre-training set size used for pre-training the CCC layer were much larger than those used for training the classification layer, we were able to use much larger hyperparameters in the CCC layer without the risk of overfitting.

Model validation. To validate the performance of our behavior classifier, we used a naive, balanced test set of behavior sequences (for definitions of common machine learning terms, see **Supplementary Table 2**). Prior to model training, we randomly selected each unique behavior sequence (or "instance") from our annotated dataset while preserving class balance. Importantly, to prevent misleading or skewed model performance results, we did not use the instances in this test set during any form of model training or adjustment.

From this balanced test set, we randomly sampled 1,000 sequences with a maximum length of 15 frames and randomly truncated 5% of these samples prior to inference to account for "edge effects," or model performance of features at the beginning or end of videos. After we used our model to infer all sampled sequences, we calculated precision, recall, F1 score, specificity, and balanced accuracy using the *sklearn.metrics* library in Python. We also calculated the normalized Matthews correlation coefficient (nMCC) using a custom Python implementation (Chicco and Jurman, 2020). We repeated this random sampling for a total of ten iterations before calculating the mean and standard deviations of each metric.

To cross validate our A-LSTM sequence model with the DEG TGM sequence model and human annotators, we first trained a TGM sequence model with a temporal window of 15 frames on the equivalent training set to that of our A-LSTM model. We then repeated the sampling and metric calculation detailed above to determine means and standard deviations for the TGM model metrics. Using both models' output prediction probabilities, we calculated the precision-recall curves for each classifier. To determine if the change between the area under the precision-recall curves (AUPRC) was significant, we used a Python version of a bootstrapping method originally implemented in R to create a normal distribution of area differences between random subsamples of the two curves with 10,000 sampling iterations (Zobolas, 2020). We then compared the area difference of the two total precision-recall curves to the mean and standard deviation of this distribution to determine significance.

Finally, we designed a custom GUI in Python that allows human annotators to classify 10,000 randomly sampled 15 frame sequences of video frames from our test set by replaying the sequence until it is classified as a behavior. Importantly, the GUI does not give the human annotator performance feedback over the course of the annotation so (much like our machine learning models) they are unable to learn as they annotate. Using these annotations, we repeated the sampling and metric calculations to determine means and standard deviations for human labeler metrics. We also incorporated this GUI into the CBAS package.

Automated inference. At the beginning of a recording, our software's inference thread creates one or more subprocesses to handle the inference of videos with DEG and DLC models. We create these subprocesses in unique Anaconda environments so both DEG and DLC models can run within their respective incompatible environments. Upon receiving the signal of new videos from the storage thread, our software's inference thread signals the active models one at a time and allows them to inference the newly collected video data. In this manner, CBAS continuously automates model inference until the recording is terminated by the user. If an A-LSTM sequence model is trained and ready for use in inference, CBAS also coordinates the automated inference of the DEG feature extractor outputs into sequenced behavior classes.

Analysis. We produced behavior actograms by binning the number of frames predicted as a given behavior over a 30 min period. For visualization, we normalized actograms to the maximum bin per recording and for each individual behavior. We produced "population" actograms by overlaying multiple individual actograms with a reduced alpha value (transparency) equal to the number of individual actograms plotted. For example, if we used CBAS to plot an actogram from a single individual mouse, it would be completely opaque, but if we plotted a population actogram from five individual mice, each actogram layer would have an alpha value of 0.2. To account for differing estrous states in female mice, we shifted their behavior actograms such that their projected day of proestrus (as determined by estrous scoring) was aligned for each mouse that we recorded. In LD, adjustments and group ns were: 0 d (n = 8 mice), -1 d (n = 5 mice) , -2 d (n = 7 mice), and -3 d (n = 7 mice). In DD, adjustments and group ns were: 0 d (n = 5 mice), -1 d (n = 4 mice), -2 d (n = 6 mice), -3 d (n = 13 mice).

To calculate circadian parameters (phase, period, and amplitude), we used CBAS to export each actogram as an .awd file, a file format compatible with ClockLab Analysis (Actimetrics), a widely-used circadian analysis software. To calculate phase, we identified activity onsets and activity offsets by using a standard template matching algorithm that searches for a 12 h period of inactivity followed by a 12 h period of activity (for onsets; the reverse was used to calculate offsets).

Onsets and offsets were manually corrected if needed (for instance, the algorithm can sometimes miss the appropriate offset if there are short bouts of activity prior to the main bout of activity). If an onset or offset occurred when the recording was briefly paused between LD and DD, we extrapolated the phase markers on just that day forwards or backwards from previous or subsequent days. To calculate period, we used a Lomb-Scargle periodogram with a range of 20 to 28 hours and a significance level of 0.001. To calculate amplitude, we measured the peak-to-peak amplitude of a sine wave fitted to the average activity profile calculated across all days in LD or all days in DD.

We performed the following statistical tests in Prism 10.0 (Graphpad): two-way ANOVA, two-way repeated measures ANOVA, mixed-effects analysis, post-hoc Tukey's multiple comparisons test. We performed a bootstrapping test in Python (Zobolas, 2020). Because no phase markers occurred at or near the 24 h modulus, we performed statistical comparisons without using circular statistics. We used Shapiro-Wilk and Brown-Forsythe tests to test for normality and equal variance, defined α as 0.05, and presented all data as mean±SEM.

Data availability

All data generated in this study that support our findings are presented within this paper or its Supplementary Materials or on the Jones lab Github page at https://github.com/jones-lab-tamu. CBAS is also available to the public at the Jones lab Github page.

Acknowledgments

We thank the members of the Jones lab for discussion and comments on the manuscript. This work was supported by National Institutes of Health Grant R35GM151020 (J.R.J.) and a Research Grant from the Whitehall Foundation (J.R.J.).

References

1. Adamah-Biassi EB, Hudson RL, Dubocovich ML (2014) Genetic deletion of MT1 melatonin receptors alters spontaneous behavioral rhythms in male and female C57BL/6 mice. Horm Behav 66:619–627.
2. Adamah-Biassi EB, Stepien I, Hudson RL, Dubocovich ML (2013) Automated video analysis system reveals distinct diurnal behaviors in C57BL/6 and C3H/HeN mice. Behav Brain Res 243:306–312.
3. Albers HE, Gerall AA, Axelson JF (1981) Effect of reproductive state on circadian periodicity in the rat. Physiol Behav 26:21–25.
4. Anderson ST, Meng H, Brooks TG, Tang SY, Lordan R, Sengupta A, Nayak S, Mřela A, Sarantopoulou D, Lahens NF, Weljie A, Grant GR, Bushman FD, FitzGerald GA (2023) Sexual dimorphism in the response to chronic circadian misalignment on a high-fat diet. Sci Transl Med 15:eabo2022.
5. Atkinson HC, Waddell BJ (1997) Circadian variation in basal plasma corticosterone and adrenocorticotropin in the rat: sexual dimorphism and changes across the estrous cycle. Endocrinology 138:3842–3848.
6. Barsha G, Denton KM, Mirabito Colafella KM (2016) Sex- and age-related differences in arterial pressure and albuminuria in mice. Biol Sex Differ 7:57.
7. Bohnslav JP, Wimalasena NK, Clausing KJ, Dai YY, Yarmolinsky DA, Cruz T, Kashlan AD, Chiappe ME, Orefice LL, Woolf CJ, Harvey CD (2021) DeepEthogram, a machine learning pipeline for supervised behavior classification from raw pixels. Elife 10 Available at: http://dx.doi.org/10.7554/eLife.63377.
8. Brzezinski D, Stefanowski J, Susmaga R, Szczech I (2020) On the Dynamics of Classification Measures for Imbalanced and Streaming Data. IEEE Trans Neural Netw Learn Syst 31:2868–2878.
9. Byers SL, Wiles MV, Dunn SL, Taft RA (2012) Mouse estrous cycle identification tool and images. PLoS One 7:e35538.
10. Chicco D, Jurman G (2020) The advantages of the Matthews correlation coefficient (MCC) over F1 score and accuracy in binary classification evaluation. BMC Genomics 21:6.
11. Fujita M, Hagino Y, Takeda T, Kasai S, Tanaka M, Takamatsu Y, Kobayashi K, Ikeda K (2017) Light/dark phase-dependent spontaneous activity is maintained in dopamine-deficient mice. Mol Brain 10:49.
12. Garner J (2017) Mouse Ethogram. Available at: https://mousebehavior.org/ethogram/ [Accessed August 17, 2022].
13. Gaskill BN, Rohr SA, Pajor EA, Lucas JR, Garner JP (2009) Some like it hot: Mouse temperature preferences in laboratory housing. Appl Anim Behav Sci 116:279–285.
14. Goulding EH, Schenk AK, Juneja P, MacKay AW, Wade JM, Tecott LH (2008) A robust automated system elucidates mouse home cage behavioral structure. Proc Natl Acad Sci U S A 105:20575–20582.
15. Grandini M, Bagli E, Visani G (2020) Metrics for Multi-Class Classification: an Overview. arXiv [statML] Available at: http://arxiv.org/abs/2008.05756.
16. Grieco F et al. (2021) Measuring Behavior in the Home Cage: Study Design, Applications, Challenges, and Perspectives. Front Behav Neurosci 15:735387.
17. Griffin AC, Whitacre CC (1991) Sex and strain differences in the circadian rhythm fluctuation of endocrine and immune function in the rat: implications for rodent models of autoimmune disease. J Neuroimmunol 35:53–64.
18. Hsu AI, Yttri EA (2021) B-SOiD, an open-source unsupervised algorithm for identification and fast prediction of behaviors. Nat Commun 12:5188.
19. Jhuang H, Garrote E, Mutch J, Yu X, Khilnani V, Poggio T, Steele AD, Serre T (2010) Automated home-cage behavioural phenotyping of mice. Nat Commun 1:68.
20. Joye DAM, Evans JA (2022) Sex differences in daily timekeeping and circadian clock circuits. Semin Cell Dev Biol 126:45–55.
21. Jud C, Schmutz I, Hampp G, Oster H, Albrecht U (2005) A guideline for analyzing circadian wheel-running behavior in rodents under different lighting conditions. Biol Proced Online 7:101–116.
22. Kahnau P et al. (2023) A systematic review of the development and application of home cage monitoring in laboratory mice and rats. BMC Biol 21:256.
23. Krizo JA, Mintz EM (2014) Sex differences in behavioral circadian rhythms in laboratory rodents. Front Endocrinol 5:234.
24. Kuljis DA, Gad L, Loh DH, MacDowell Kaswan Z, Hitchcock ON, Ghiani CA, Colwell CS (2016) Sex Differences in Circadian Dysfunction in the BACHD Mouse Model of Huntington's Disease. PLoS One 11:e0147583.
25. Kuljis DA, Loh DH, Truong D, Vosko AM, Ong ML, McClusky R, Arnold AP, Colwell CS (2013) Gonadal- and sex-chromosome-dependent sex differences in the circadian system. Endocrinology 154:1501–1512.
26. Lee TM, Hummer DL, Jechura TJ, Mahoney MM (2004) Pubertal development of sex differences in circadian function: an animal model. Ann N Y Acad Sci 1021:262–275.
27. Li J-D, Hu W-P, Boehmer L, Cheng MY, Lee AG, Jilek A, Siegel JM, Zhou Q-Y (2006) Attenuated circadian rhythms in mice lacking the prokineticin 2 gene. J Neurosci 26:11615–11623.
28. Mathis A, Mamidanna P, Cury KM, Abe T, Murthy VN, Mathis MW, Bethge M (2018) DeepLabCut: markerless pose estimation of user-defined body parts with deep learning. Nat Neurosci 21:1281–1289.
29. Merchenthaler I, Lane MV, Numan S, Dellovade TL (2004) Distribution of estrogen receptor alpha and beta in the mouse central nervous system: in vivo autoradiographic and immunocytochemical analyses. J Comp Neurol 473:270–291.
30. Metzger J, Wicht H, Korf H-W, Pfeffer M (2020) Seasonal Variations of Locomotor Activity Rhythms in Melatonin-Proficient and -Deficient Mice under Seminatural Outdoor Conditions. J Biol Rhythms 35:58–71.
31. Muller M, Wolf CT, Andres J, Desmond M, Joshi NN, Ashktorab Z, Sharma A, Brimijoin K, Pan Q, Duesterwald E, Dugan C (2021) Designing Ground Truth and the Social Life of Labels. In: Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems, pp 1–16 CHI '21. New York, NY, USA: Association for Computing Machinery.

32. Nilsson SRO, Goodwin NL, Choong JJ, Hwang S, Wright HR, Norville ZC, Tong X, Lin D, Bentzley BS, Eshel N, McLaughlin RJ, Golden SA (2020) Simple Behavioral Analysis (SimBA) – an open source toolkit for computer classification of complex social behaviors in experimental animals. bioRxiv:2020.04.19.049452 Available at: https://www.biorxiv.org/content/10.1101/2020.04.19.049452v2 [Accessed February 22, 2024].
33. Pasquali V, Capasso A, Renzi P (2010) Circadian and ultradian rhythms in locomotory activity of inbred strains of mice. Biol Rhythm Res 41:63–74.
34. Pendergast JS, Branecky KL, Yang W, Ellacott KLJ, Niswender KD, Yamazaki S (2013) High-fat diet acutely affects circadian organisation and eating behavior. Eur J Neurosci 37:1350–1356.
35. Pereira TD, Shaevitz JW, Murthy M (2020) Quantifying behavior to understand the brain. Nat Neurosci 23:1537–1549.
36. Peters SM, Pothuizen HHJ, Spruijt BM (2015) Ethological concepts enhance the translational value of animal models. Eur J Pharmacol 759:42–50.
37. Richardson CA (2015) The power of automated behavioural homecage technologies in characterizing disease progression in laboratory mice: A review. Appl Anim Behav Sci 163:19–27.
38. Robinson-Junker AL, O'hara BF, Gaskill BN (2018) Out Like a Light? The Effects of a Diurnal Husbandry Schedule on Mouse Sleep and Behavior. J Am Assoc Lab Anim Sci 57:124–133.
39. Royston SE, Yasui N, Kondilis AG, Lord SV, Katzenellenbogen JA, Mahoney MM (2014) ESR1 and ESR2 differentially regulate daily and circadian activity rhythms in female mice. Endocrinology 155:2613–2623.
40. Salem GH, Dennis JU, Krynitsky J, Garmendia-Cedillos M, Swaroop K, Malley JD, Pajevic S, Abuhatzira L, Bustin M, Gillet J-P, Gottesman MM, Mitchell JB, Pohida TJ (2015) SCORHE: a novel and practical approach to video monitoring of laboratory mice housed in vivarium cage racks. Behav Res Methods 47:235–250.
41. Sanchez-Alavez M, Alboni S, Conti B (2011) Sex- and age-specific differences in core body temperature of C57Bl/6 mice. Age 33:89–99.
42. Schwartz WJ, Zimmerman P (1990) Circadian timekeeping in BALB/c and C57BL/6 inbred mouse strains. J Neurosci 10:3685–3694.
43. Segalin C, Williams J, Karigo T, Hui M, Zelikowsky M, Sun JJ, Perona P, Anderson DJ, Kennedy A (2020) The Mouse Action Recognition System (MARS): a software pipeline for automated analysis of social behaviors in mice. Cold Spring Harbor Laboratory:2020.07.26.222299 Available at: https://www.biorxiv.org/content/10.1101/2020.07.26.222299v1 [Accessed December 7, 2020].
44. Shuboni-Mulligan DD, Young DL Jr, De La Cruz Minyety J, Vera E, Munasinghe J, Gall AJ, Gilbert MR, Armstrong TS, Smart DK (2021) Impact of age on the circadian visual system and the sleep-wake cycle in mus musculus. NPJ Aging Mech Dis 7:10.
45. Siepka SM, Takahashi JS (2005) Methods to Record Circadian Rhythm Wheel Running Activity in Mice. In: Methods in Enzymology (Young MW, ed), pp 230–239. Academic Press.
46. Sotelo MI, Tyan J, Markunas C, Sulaman BA, Horwitz L, Lee H, Morrow JG, Rothschild G, Duan B, Eban-Rothschild A (2022) Lateral hypothalamic neuronal ensembles regulate pre-sleep nest-building behavior. Curr Biol Available at: http://dx.doi.org/10.1016/j.cub.2021.12.053.
47. Starnes AN, Jones JR (2023) Inputs and Outputs of the Mammalian Circadian Clock. Biology 12:508.
48. Staudemeyer RC, Morris ER (2019) Understanding LSTM – a tutorial into Long Short-Term Memory Recurrent Neural Networks. arXiv [csNE] Available at: http://arxiv.org/abs/1909.09586.
49. Steele AD, Jackson WS, King OD, Lindquist S (2007) The power of automated high-resolution behavior analysis revealed by its application to mouse models of Huntington's and prion diseases. Proc Natl Acad Sci U S A 104:1983–1988.
50. Ström JO, Theodorsson A, Ingberg E, Isaksson I-M, Theodorsson E (2012) Ovariectomy and 17β-estradiol replacement in rats and mice: a visual demonstration. J Vis Exp:e4013.
51. Tillmann JF, Hsu AI, Schwarz MK, Yttri EA (2024) A-SOiD, an active-learning platform for expert-guided, data-efficient discovery of behavior. Nat Methods Available at: http://dx.doi.org/10.1038/s41592-024-02200-1.
52. van der Veen R, Abrous DN, de Kloet ER, Piazza PV, Koehl M (2008) Impact of intra- and interstrain cross-fostering on mouse maternal care. Genes Brain Behav 7:184–192.
53. van Oosterhout F, Lucassen EA, Houben T, vanderLeest HT, Antle MC, Meijer JH (2012) Amplitude of the SCN clock enhanced by the behavioral activity rhythm. PLoS One 7:e39693.
54. Vaswani A, Shazeer N, Parmar N, Uszkoreit J, Jones L, Gomez AN, Kaiser L, Polosukhin I (2017) Attention Is All You Need. arXiv [csCL] Available at: http://arxiv.org/abs/1706.03762.
55. Verwey M, Robinson B, Amir S (2013) Recording and analysis of circadian rhythms in running-wheel activity in rodents. J Vis Exp Available at: http://dx.doi.org/10.3791/50186.
56. Walton JC, Bumgarner JR, Nelson RJ (2022) Sex differences in circadian rhythms. Cold Spring Harb Perspect Biol 14 Available at: https://cshperspectives.cshlp.org/content/early/2022/01/31/cshperspect.a039107.short?casa_token=o8jmv70ohpsAAAAA:PRsQUgDmTXWRoMV1VShAAsMPEHyzVKZhhxTbcLPrbc8i61ioEFq40E6_CoCWdFC6y89C6RCr.
57. Xie S, Sun C, Huang J, Tu Z, Murphy K (2017) Rethinking Spatiotemporal Feature Learning: Speed-Accuracy Trade-offs in Video Classification. arXiv [csCV] Available at: http://arxiv.org/abs/1712.04851.
58. Yamanaka Y, Honma S, Honma K-I (2013) Daily exposure to a running wheel entrains circadian rhythms in mice in parallel with development of an increase in spontaneous movement prior to running-wheel access. Am J Physiol Regul Integr Comp Physiol 305:R1367–R1375.
59. Zhang C, Li H, Han R (2020) An open-source video tracking system for mouse locomotor activity analysis. BMC Res Notes 13:48.
60. Zobolas J (2020) usefun. https://github.com/bblodfon/usefun Available at: doi.org/10.5281/zenodo.10694717 [Accessed February 22, 2024].