Internship

3P platform

2024/2025

Santiago Holguin Urbano

Identification of major behavioral differences across life in individuals exposed to different types of maternal neglect

Encadrant: Marion Rivalan

Paris-Saclay institute of neurosciences (NeuroPSI)

Mail: marion.rivalan@cnrs.fr

Tel : 01 69 82 46 78

151 route de la Rotonde CEA
Paris-Saclay, Bat.151

Tab of content

Abbreviations list

US - United States
ADHD – Attention Deficit Hyperactivity Disorder
AMBER – Automated Maternal Behavior during Early Life in Rodents
DLC – DeepLabCut
ELA – Early Life Adversity
HPA Axis – Hypothalamic-Pituitary-Adrenal Axis
LBN – Limited Bedding and Nesting Model
NMDAR – N-Methyl-D-Aspartate Receptor
PFC – Prefrontal Cortex
PV – Parvalbumin (a type of interneuron)
SimBA – Simple Behavioral Analysis
Deg - DeepEthogram
PCA - Principal component analysis
CNNs - Convolutional neural networks
PELT - Pruned Exact Linear Time

Keywords

Maternal neglect
Machine learning methods in behavioural neuroscience
Limited bedding and nesting material
Mice model
Early life adversity/stress

Acknowledgments

First of all, I would like to warmly thank Marion Rivalan for her guidance, availability, and support throughout my internship. I would also like to thank Sylvie Granon for welcoming me into her team, where I was very well received.
I would like to thank Laure Lemercier for her help in explaining all parts of the project to me, as well as for her excellent work in annotating maternal videos and collecting pup's data.
Finally, a big thank you to the whole team for being kind throughout my internship.

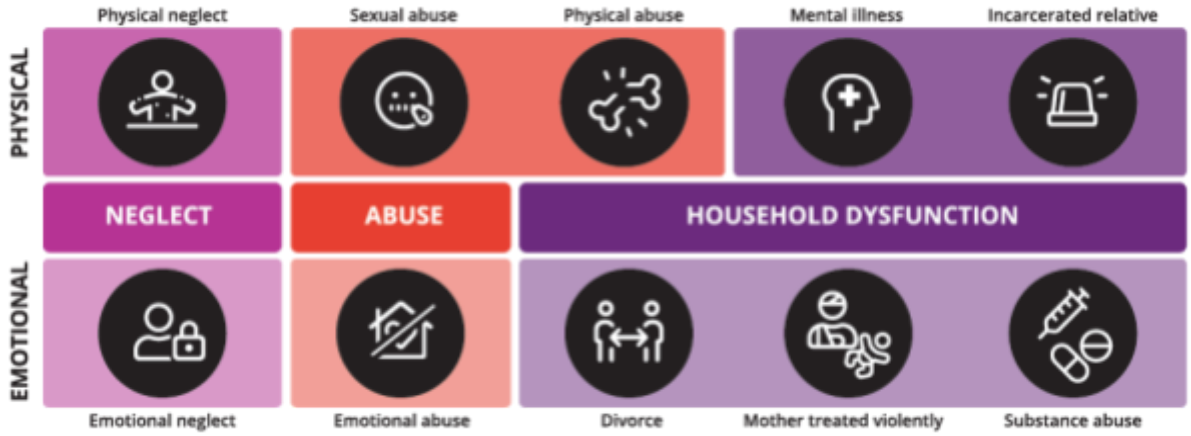


Figure 1. Types of child maltreatment and neglect. Emotional and physical forms of maltreatment are illustrated (Ayre et Krishnamoorthy 2020).

Introduction

A. Child abuse and maternal neglect

Child abuse remains a significant issue in today's society, affecting over 100 million children between 1997 and 2015 who were victims of different forms (Figure 1. of maltreatment in the European region (WHO et Bring 2015). It's defined as all forms of physical and emotional ill-treatment, and exploitation that results in actual or potential harm to the child's health (WHO et S.Becker 2024). Among these forms, one of the most widespread is neglect (Figure 1). It's defined as inadequate health care, education, supervision, protection from hazards in the environment, and unmet basic needs such as clothing and food. It includes cognitive, supervision, emotional and physical neglect.

The severe short and long-term consequences of child abuse emphasize the need for a deeper understanding of this phenomenon, as such knowledge can help improve the lives of affected individuals and contribute to its prevention. Nevertheless, research on child abuse in humans presents significant challenges due to limited information from victims and parents, as well as the difficulty of tracking the long-term effects of neglect from its onset.

To overcome these challenges, animal models can be used. Rodents, in particular, serve as powerful models for studying mammalian brain development due to their structural and functional similarities to the human nervous system. Their use allows researchers to isolate specific neurobiological mechanisms in a controlled setting, free from human cultural influences (Bryda 2013).

B. Mice as a model of maternal neglect - LBN

Rodents (and primates) have been used as models to study early life adversity (ELA), a term referring to a significant period during child experiments of trauma or social deprivation during early postnatal life (Veenema 2009). Those studies brought new understanding about ELA consequences, as for example the dysregulation of key neurobiological systems, including vasopressin, serotonin, and the hypothalamic– pituitary–adrenal (HPA) axis (Haller et al. 2014).

Some rodent models commonly used include : maternal separation (3 hours per day during the first two weeks of life), post-weaning social isolation (only visual auditory and olfactory contact), and post-weaning social deprivation (no interaction, or interaction of young males exclusively with females).

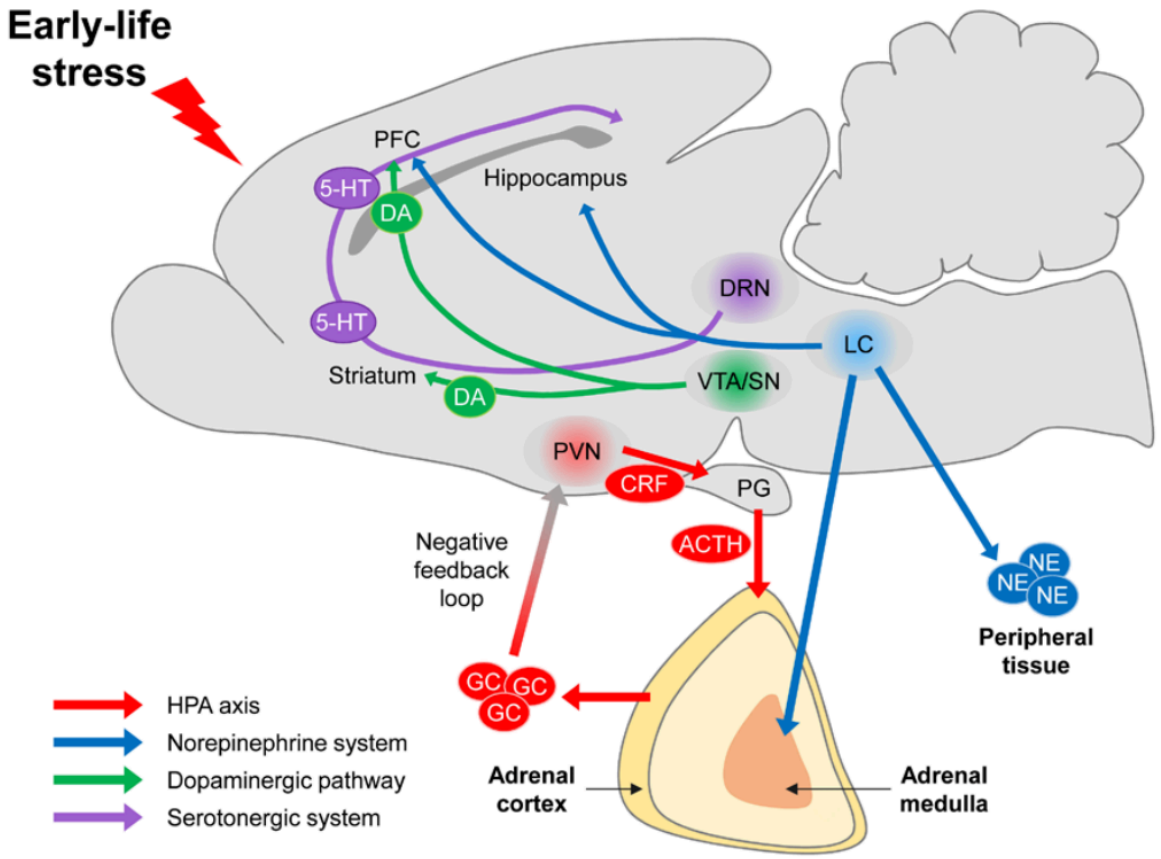


Figure 2. Neurobiological pathways affected by early-life stress. Early-life stress activates the hypothalamic-pituitary-adrenal (HPA) axis (red arrows), beginning with CRF release from the PVN, followed by ACTH from the pituitary gland (PG), and glucocorticoid (GC) secretion from the adrenal cortex, which exerts negative feedback. Stress also modulates the serotonergic system (purple arrows) from the dorsal raphe nucleus (DRN) to the prefrontal cortex (PFC), hippocampus, and striatum. Dopaminergic pathways (green arrows) from the ventral tegmental area/substantia nigra (VTA/SN) project to the PFC and striatum. The noradrenergic system (blue arrows), originating in the locus coeruleus (LC), affects both central and peripheral targets. These interconnected systems contribute to stress regulation and brain development (Lee et Jung 2024).

Most maternal separation models involve temporary separation and abandonment of pups; and do not fully capture key characteristics of maternal neglect, such as the mother's physical presence (Haller et al. 2014; Veenema 2009). The underlying mechanisms of neglect are better exposed on models that emphasize the quality of maternal care. In these models dam's experience stress due to harsh conditions, such as low resources and inadequate housing, inducing fragmented maternal care. The maternal neglect model we'll use follows these principles; Limited bedding and nesting model (LBN) is a neglect model where cages are fitted with a wire mesh platform, positioned some centimeters above the cage floor; bedding is reduced, covering the cage floor only sparsely, and nesting material is significantly limited (Walker et al. 2017).

C. LBN consequences on pups

The cortisol system plays a central role in regulating the stress response through the hypothalamic-pituitary-adrenal (HPA) axis (Hosseinichimeh, Rahmandad, et Wittenborn 2015). When an organism encounters a stressor (Figure 2), the hypothalamus stimulates the release of adrenocorticotropic hormone (ACTH) from the pituitary gland, which in turn triggers cortisol secretion from the adrenal glands (Papadimitriou et Priftis 2009). Cortisol acts on multiple organs, including the brain, to mobilize energy needed for the stress response and modulate physiological and behavioral functions.

In the brain, cortisol primarily binds to two types of receptors, glucocorticoid receptors (GR) and mineralocorticoid receptors (MR), which regulate the transcription of genes involved in neuronal plasticity, stress adaptation, and neurodevelopmental processes (Herman 2022). Maintaining balanced cortisol levels is crucial for healthy brain development. Both elevated (hypercortisolism) and reduced (hypocortisolism) cortisol levels can negatively affect the maturation of the brain and the development of emotional and cognitive functions (Lautarescu, Craig, et Glover 2020).

Pup's exposed to fragmented maternal care in LBN conditions exhibit decreased expression of glucocorticoid and mineralocorticoid receptors, altered basal corticosterone levels and genetic dysregulation of stress response genes (Bath, Manzano-Nieves, et Goodwill 2016). Therefore LBN modifies the normal trajectory of pup's neurodevelopment by altering dam behaviours (Duffy, McLaughlin, et Green 2018).

D. Maternal care in mice

LBN model differ from other ELA models by modifying the mother (and pup's) environment, inducing stressful conditions of home cage living arrangement for the dam and leading to a disruption of normal maternal care.

Parental care has been divided in different phases starting by the nest building, then incubation, and both feeding and protection of young. Indeed, maternal behaviours start during early gestation, by the preparation of a brood nest : a completely enclosed nest. This nest acts as a sleeping and thermoregulator litter, and is normally 2-3 times bigger than the dams (Lynch et Possidente 1978).

Nursing is the most frequent maternal behavior in mice, accounting for approximately 92% of maternal care during the first three weeks postpartum (König et Markl 1987). Initially, the mother remains curled around the pups in the nest, providing warmth and milk. By day 9, the dam begins engaging in other maternal behaviors, such as licking and grooming the pups. When a pup is displaced, the mother quickly retrieves it, responding to vocal and olfactory cues. This Pup retrieval response is essential for maintaining the litter's cohesion and protection (Weber et Olsson 2008).

E. Maternal care in LBN mice

These normal behaviors ensure an optimal environment for pup development and survival. However, stress conditions can significantly disrupt these patterns, affecting pup health, as previously discussed. LBN model induces fragmented and unpredictable maternal care: dams spend more time on the nest, but they also show increased nest entry and exit frequency. Additionally, LBN is associated with increased dam-pup interactions and the potential emergence of abuse-like behaviors, such as forceful kicking of pups. LBN dams also exhibit fragmentation in feeding behavior, characterized by shorter but more frequent eating periods. Variations in licking and grooming patterns have been observed, with LBN dams sometimes engaging in excessive grooming and, at other times, neglecting their pups (Gallo et al. 2019).

Delving deeper, LBN dams exhibit increased self-grooming, likely as a stress-coping mechanism. Instead of methodically moving pups, some dams engage in aggressive-like carrying or even display neglect. The most significant differences in maternal behavior occur during transitions between night and day. At these moments, LBN dams display more erratic behaviors, rapidly switching between eating, grooming, and tending to pups, whereas control dams maintain a more structured caregiving approach (Demaestri et al. 2022).

Previous studies have often examined the behaviors of pups and their dams independently, rather than considering them as part of a dynamic, interconnected system. One major limitation of this dual approach is the difficulty in accurately classifying behaviors – especially when relying on manual annotation, which is labor-intensive and time-consuming. This methodological challenge hinders researchers from capturing a comprehensive picture of maternal care and pup responses within the same study, limiting insights into how their interactions jointly influence developmental and health outcomes.

F. Maternal care behaviours annotation

Manual scoring of maternal behaviors depends on trained observers and is both time-consuming and susceptible to inter-rater variability. Although this method is generally considered highly reliable, its accuracy can be affected as the volume of data increases : either by reducing observer efficiency or requiring more observers. As a result, while manual scoring offers strong reliability, it lacks scalability for large datasets. On the other hand, automated behavior classification improves scalability, precision, and objectivity (Worley, Djerdjaj, et Christianson 2019).

In biology and behavioral studies, DeepLabCut (DLC) has been widely used to track various body parts in animal videos (Sturman et al. 2020). Automated Maternal Behavior during Early Life in Rodents (AMBER), is one of the pre-trained pipelines that integrates DLC for pose tracking to then identify maternal behaviors. DLC tracks 32 points on rat dams and 9 points on pups, providing detailed movement data. These coordinates are then processed by a pre-trained Random Forest model of Simple Behavioral Analysis (SimBA), which classifies different behaviours as nursing, passive nursing, nest attendance, licking and grooming, self-directed grooming, eating, and drinking (Lapp, Salazar, et Champagne 2023).

In personal communication with the author of AMBER, it was confirmed that AMBER currently performs poorly when estimating mouse behaviors instead of rat behaviors. Therefore, as an alternative we used DeepEthogram (Deg), another promising model for automated annotation of behaviour but which classifies behaviors directly from the raw pixel values of videos (Bohnslav et al. 2021).

G. Objectives and scientific questions

The **primary objective** of this study was to identify which is the most suitable machine learning method for accurately detecting specific maternal behaviors within a homecage. As a **secondary objective**, this study also seeks to deepen our understanding of maternal neglect within the LBN model. Specifically, we aim to investigate the outcomes associated with early-life adversity. To this end, the study is structured around the already collected data on motor, transcriptomic, hormonal and physical parameters of LBN pups at juvenile stage. And the specific maternal behaviors and their evolution throughout the postnatal period in LBN, and control conditions. The **third objective** is to investigate intra-group variability among pups, aiming to identify whether certain individuals exhibit similar phenotypes and cluster together. This analysis will explore whether the similarity in pup phenotypes is influenced by the Dams specificities they were raised with, and therefore by the maternal behaviors they experienced —or failed to experience— under LBN conditions.

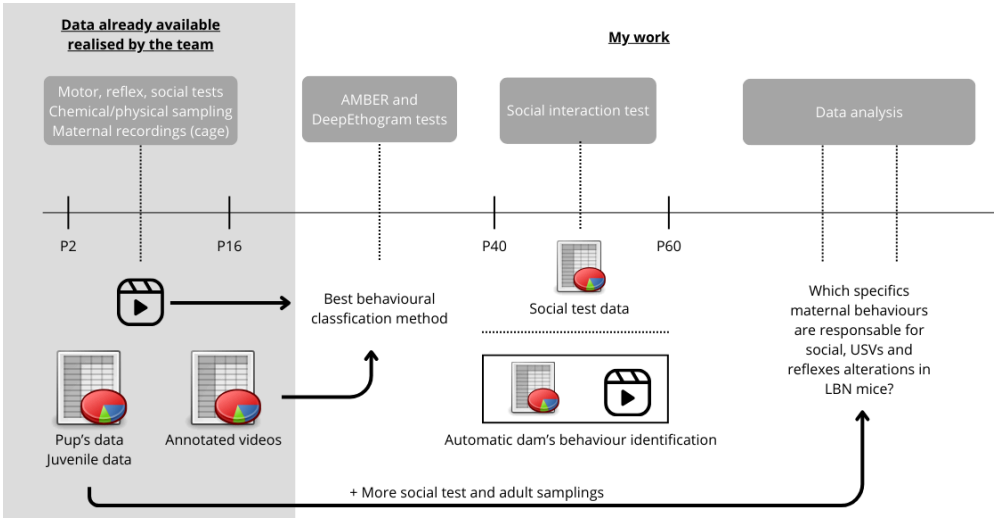


Figure 3. Methodological and experimental approaches. Data and experiments already done by the team are in the grey rectangle.

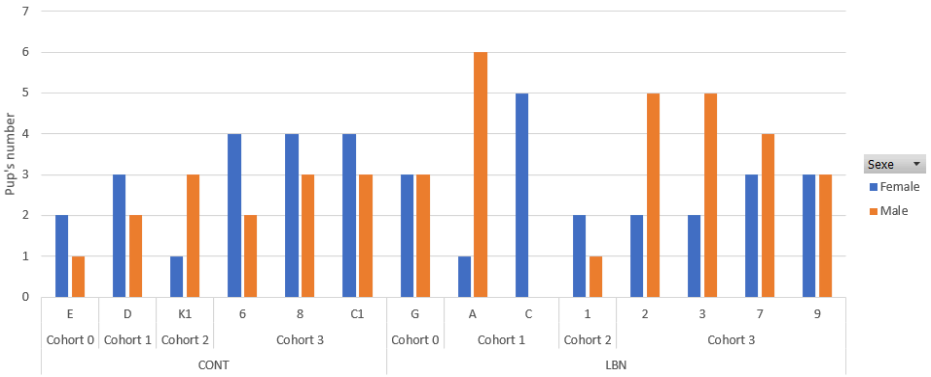


Figure 4. LBN and control cages across cohorts. The study was conducted in multiple repetitions, using distinct cohorts. Each cohort included a combination of LBN and control cages

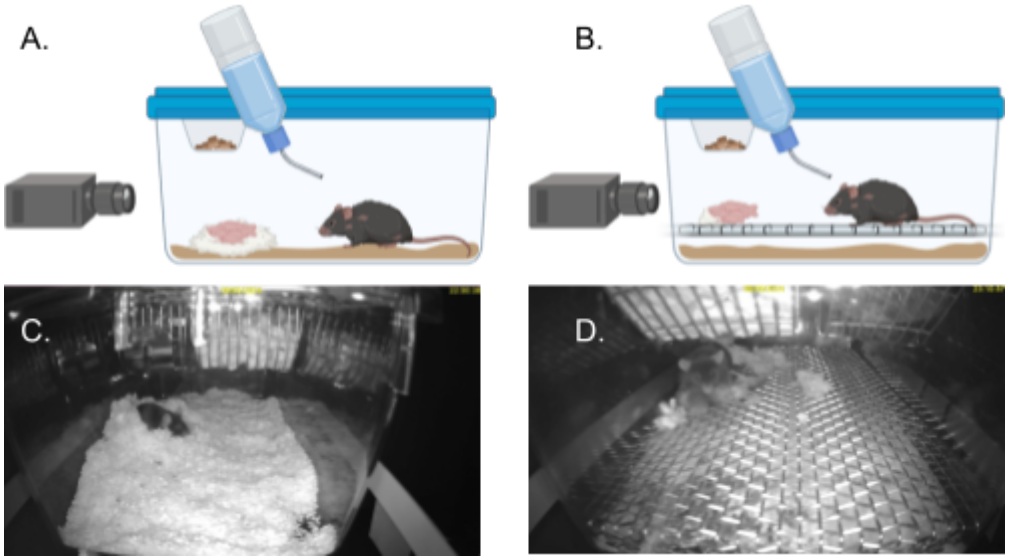


Figure 5. LBN setup for recording. In normal conditions (A) dam and pups are directly on the ground having access to bedding, a piece of cotton of 4x4 cm is accessible and serves as nesting material. In LBN conditions (B), dams and pups are placed on an elevated wire mesh platform reducing the access to bedding material, and identification material is reduced by 50% (2 x 4 cm cotton piece). Corresponding recorded examples are shown for control conditions (C) and for LBN condition (D).

Materials and methods

Over the past few months, the team has been collecting both maternal data by recording maternal behaviours and their interactions with pups (Figure 3). In addition, the team has tested pups' reflexes, collected blood samples, recorded pup's weights, and, for some cohorts, extracted brain tissue for neurotransmitter analyses. Maternal data will be analysed using deep learning methods and will be co-analysed not only with existing pup data but also with new data obtained through social and motor tests conducted during my internship (Figure 3).

Animals

This study was conducted using wild-type C57BL/6N mice. A specific strain with well-characterized genetic background and stable maternal behaviors. The experimental design included a total of 8 dams and 48 pups subjected to the Limited Bedding and Nesting (LBN) model, while 6 dams and 32 pups were used as controls (Figure 4). This design ensures a sufficient sample size for statistical analyses, allowing both parametric and non-parametric tests to compare LBN-exposed and control mice effectively. Both male and female pups were included in the study, with a distribution of 21 LBN females, 27 LBN males, 18 control females, and 14 control males in total (Figure 4).

Recordings

To achieve optimal visualization of maternal and pup interactions, cameras were positioned in a front configuration, allowing multi-angle tracking of behavior (Figure 5.C-D). This setup was recommended by the AMBER pipeline's authors to improve pose estimations.

For high-resolution behavioral analysis, multiple videos were recorded per cage at two different periods of the day : during the first 4 night hours and during the 4 last night hours. Each recording lasted 1 minute, with recordings repeated every 10 minutes to capture maternal interactions and pup behaviors throughout two different periods. Control mice followed the same recording pattern. A first cohort was fully labeled manually by a human experimenter and has been used as a reference dataset for evaluating and training Deg models. Since a progressive lighting change in the room starts at 07:30, only 3 and a half hours of first phase recording are used.LBN and control cages from cohort 0 will not be used for results analysis because their recordings were done from side view and not front view.

LBN Model

The home-cages of both control and LBN mice are of the same size maintaining a standardized rearing environment across all experimental groups. The maternal behaviors were recorded from postnatal day P4 to P11 (Figure 5), ensuring that the analysis covered early postnatal development.

Initially, dams and pups were housed in a standard, richly furnished cage, ensuring that maternal care was not disrupted at birth. After 3 days, they were transferred to an impoverished cage environment, designed to induce maternal stress and disrupt stable caregiving behaviors (Figure 5.B). The LBN condition was maintained for 7 days, allowing to assess both acute and long-term effects of fragmented maternal care on pup neurodevelopment. The choice of late environmental inmersion and short time maintenance of LBN condition aims to ameliorate pup survey, and to reduce too strong effects on dams and pups.

The key modifications in the LBN model included:

- Metallic Grid: A wire-mesh platform was placed some cm above the floor, preventing the accumulation of nesting material and separating this material from the dam.
- Reduced Bedding Material: Nesting material was reduced by 50%, limiting the dam's ability to construct a thermally stable nest and increasing pup and dams exposure to stress conditions such as inconsistent warmth.

By manipulating maternal resource availability, the LBN model allows us to analyze how environmental stress shapes maternal behavior and pup neurodevelopment.

Automatic Classification of Maternal Behaviors

AMBER

The AMBER pipeline has been tested as the primary tool for classifying key maternal behaviors. AMBER was originally **trained on rat** datasets (Lapp, Salazar, et Champagne 2023). Recently discussing directly with the authors of AMBER, they shared with us their new mice model which was not trained on B6 mice but only trained with CD1 mice, white and bigger mice; additionally, recordings were taken from side view camera configuration. The AMBER pipeline performance has been tested using different maternal videos from the first cohort.

This pipeline uses pre-trained weights to make two resnet-50 DLC models adapted to labelise on one side 32 maternal key body parts, and on another side 9 pup key body parts. Both maternal and pup estimations are then merged and used by 7 behavior classifiers (Random forest models) trained with SimBA to estimate 7 different behaviours.

This pipeline results in an accurate rat behaviour classification. All classifiers were described as having a strong performance, with F1 scores above 0.886 during training. On hold-out videos, performance remained high for nest attendance (F1 = 0.990), active nursing (F1 = 0.828), and licking and grooming (F1 = 0.766), while it was moderate for eating, drinking, and self-directed grooming (F1 = 0.534–0.554).

As explained by the authors of AMBER, the first step prior to behaviour classification involves validating the pose-estimation results –a particularly sensitive step due to the characteristics of the animals in our videos. Videos were selected for testing the pose-estimation pipeline based on two criteria: overall video quality and clear visibility of both dam and pup body parts.

F1

The F1 score is a commonly used metric to evaluate the performance of classification models. It is the harmonic mean of **precision (how many selected items are relevant) and recall (how many relevant items are selected)**. An F1 score of 1.0 indicates perfect precision and recall, while a score of 0 indicates poor performance. It is especially useful when dealing with imbalanced data, as it provides a balanced measure of a model's accuracy.

While accuracy measures overall correctness, it can be misleading for imbalanced datasets. For example, in a binary task with only 5% positives, a model predicting only negatives achieves 95% accuracy but detects no positives. In such cases, the F1 score—balancing precision (true positives among predicted positives) and recall (true positives among actual positives)—offers a better evaluation of the minority class, especially when correct detection is critical.

$$F_1 = 2 \cdot \frac{\mathrm{Precision} \cdot \mathrm{Recall}}{\mathrm{Precision} + \mathrm{Recall}} \qquad \mathrm{With}: \mathrm{Precision} = \frac{TP}{TP + FP} \quad \text{and} \quad \mathrm{Recall} = \frac{TP}{TP + FN}$$

Deep Ethogram (Deg)

Deep Ethogram (Bohnslav et al. 2021) is a machine learning pipeline that uses three different models to build an ethogram from raw video pixels. The first model, called optical flow, converts static pixel data from groups of frames into a new set of frames encoding spatial information and capturing displacements occurring between consecutive frames. This first step can be done in two

different ways (two different model architectures): one capturing 2D data and another capturing 3D data. The first model architecture (2D) is part of a configuration called “fast”, while the second model (3D) is part of another configuration called “slow”.

Two different convolutional neural networks (CNNs) receive either the optical flow or the static pixel data to compress the large number of raw pixels into a small set of feature vectors representing the probability of each behavior on each frame (the final probabilities are the mean of both outputs). The architecture of both CNNs can change depending on the configuration: the “fast” configuration uses the ResNet-18 architecture (18 neuron layers), and the “slow” configuration uses the ResNet-50 architecture (50 neuron layers).

A final sequence model uses longer temporal information to improve the first CNNs’ predictions (the first model uses either 1 frame or about 10–11 frames of information, less than a second), memorising last-labeled data and capturing not only motion but also contextual data.

Deg can be configured and trained to automatically label all possible behaviours. For our project we’ll use Deg to automatically annotate these behaviours :

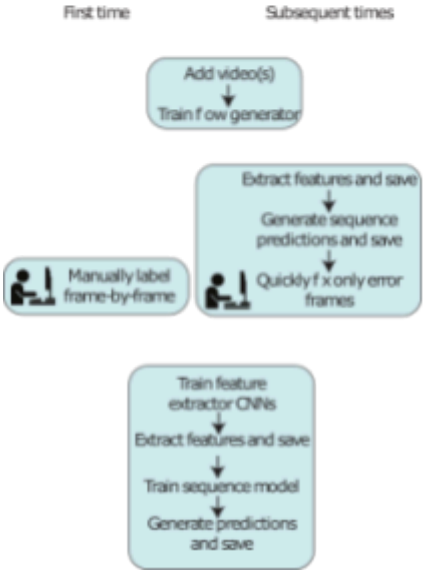
- Arched back nursing: Differences in maternal postures and pup attachment.
- Licking and grooming : Frequency and duration of pup-directed grooming.
- Pup retrieval: Retrieving displaced pups.
- Self-grooming: An indicator of maternal stress and coping strategies.
- Feeding and drinking patterns: To assess fragmentation in maternal behaviors under LBN conditions.
- Nest attendance : Two different behaviours on nest and off nest
- Nest building : Nesting material displacement
- Kicking : Aggressive kicking like behavior ; not sufficiently represented in training data for reliable classification

Aggressive interaction behaviour initially thought as present will not be well annotated because we only found-it in three videos. Deg performances are highly variable and are mostly dependent on the quantity of training data. The authors recommend 9 000 frames per behaviour to obtain a 0.7 F1.

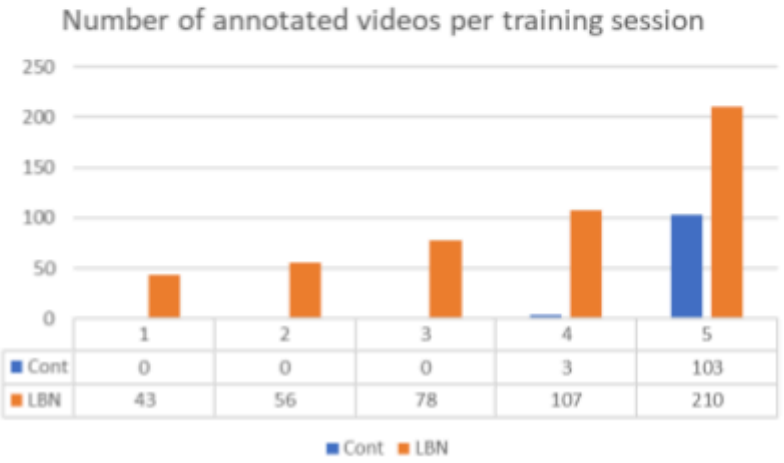
A.

Frame	Empty	Onnest	Offnest	ABN	Carryingpups	Selfgrooming	Eat_drink	Kicking	Groomingpups	Build
0	0	1	0	0	0	1	0	0	0	0
1	0	1	0	0	0	1	0	0	0	0
2	0	1	0	0	0	1	0	0	0	0
3	0	1	0	0	0	1	0	0	0	0
4	0	1	0	0	0	1	0	0	0	0
5	0	1	0	0	0	1	0	0	0	0
6	0	1	0	0	0	1	0	0	0	0
7	0	1	0	0	0	1	0	0	0	0
8	0	1	0	0	0	1	0	0	0	0
9	0	1	0	0	0	1	0	0	0	0
10	0	1	0	0	0	1	0	0	0	0
11	0	1	0	0	0	1	0	0	0	0
12	0	1	0	0	0	1	0	0	0	0
13	0	1	0	0	0	1	0	0	0	0
14	0	1	0	0	0	1	0	0	0	0
15	0	1	0	0	0	1	0	0	0	0
16	0	1	0	0	0	1	0	0	0	0
17	0	1	0	0	0	1	0	0	0	0
18	0	1	0	0	0	1	0	0	0	0
19	0	1	0	0	0	1	0	0	0	0
20	0	1	0	0	0	1	0	0	0	0
21	0	1	0	0	0	1	0	0	0	0
22	0	1	0	0	0	1	0	0	0	0
23	0	1	0	0	0	1	0	0	0	0
24	0	1	0	0	0	1	0	0	0	0
25	0	1	0	0	0	1	0	0	0	0

B.



C.



	1	2	3	4	5
Cont	0	0	0	3	103
LBN	43	56	78	107	210

Figure 6. Annotation Process and Training Sessions. (A) Example of csv file used for deg training. (B) Schematic of the recommended DEG training pipeline (adapted from the DEG GitHub), which we followed to train behavior classifiers. The pipeline includes three steps: training the flow generator, the feature extractor, and the sequence model. (C) Number of annotated videos included in each training session (TS1–TS5), with a progressive increase in training data. The bar chart indicates the number of videos per condition (Control vs LBN) included at each training step.

Training the models

Deg receives three different types of data as input to train its models. First, raw video data is used to train the **flow generator model**, without any supplementary information. Then, the **feature extraction model** requires annotated videos in the form of a binary frame-by-frame ethogram (Figure 6.A). Finally, the **sequence model** is trained using the same annotated data, plus the output from the feature extraction model.

In this context, the **training process** refers to the phase in which the model learns to recognize and classify behaviors by adjusting its internal parameters based on labeled video data. This is achieved through multiple training sessions (TS), during which the model iteratively optimizes its predictions to reduce errors on the training set. The **validation process**, on the other hand, serves as an independent assessment of the model's performance (see next section).

Our objective was to match or surpass AMBER-rat performance validation results for each of the maternal behaviour. To do so, we performed five sequential TS using the "**fast**" configuration (less performant but faster to generate outputs). Our training procedure was straightforward, following the recommended training pipeline (Figure 6.B). We **gradually increased the training dataset** with each session until achieving satisfactory F1 and accuracy scores. Videos were selected and added based on model performance needs, with the aim of improving classification of underperforming behaviors. 80% of annotated videos were used for training and 20% for validation for TS 1-6, but this ratio was changed for TS 7 to increase training data quantity.

The video selection process in the first two training sessions was only exploratory. This meant that only common behaviours were present in training because of the selection technique. TS1 included 43 videos, and TS2 56. Some **cohort 1 LBN videos** were included in TS2, but removed in TS3 (Figure 6.C) due to poor model performance probably due to the fact that the cage was inverted in this cohort LBN videos.

In TS3, we used the "findFreq.py" script to quantify the number of annotated frames per behavior and identify behaviors lacking annotations. We then selected videos containing these behaviors, increasing the training set to 107 videos and enriching the representation of previously rare behaviors. The same approach was applied in TS4, where we added 3 control videos from cohort 0 and **25 self-annotated Cohort 1 videos**, randomly selected for annotation.

TS4 yielded very good results, so for TS5 we focused on behaviors that remained poorly detected. Since all previously annotated videos had been used for training or validation, we selected **new videos for annotation and model training**. These were chosen by watching the videos and checking for the presence of those poorly detected behaviours. After selecting 88 such videos, they were manually annotated. 100 pre-annotated control videos chosen in a similar way as for TS3 were added, but this time **targeting control data** specifically equilibrating the proportion of each available behaviour.

TS6 was trained with the same videos as TS5, but with the kicking behavior excluded, and it's excluded from results because of no performance difference compared to TS5. So far, we have only fully **validated and tested the model from TS5**, which will be used from here forward for automated behavioral annotation.

For TS5, we had 500 labeled frames for "Carrying pups" behaviour. In order to **augmentate training data** without having to search and search occurrences of this behaviour, we generated synthetic variants using common video augmentation techniques (Shorten et Khoshgoftaar 2019). Those techniques included : horizontal flipping, brightness and contrast enhancement, slight zoom and cropping, and Gaussian blurring. We also created combinations of these transformations to further diversify the input space. These augmentations increased the dataset to **roughly 3,500 frames** for "Carrying pups". TS6 was TS5 but without kicking as a possible behavior, TS7 followed the same methodology as TS5 but with more LBN videos and with less not relevant control videos.

Validation of trained models

Validation is a technique used to assess how well a model generalizes to new, unseen data (here presented as validation data). In the case of DEG, validation helps ensure that the model **doesn't simply memorize the training videos but instead learns patterns** that apply to new animals, days, or cages. This is especially important when working with complex and variable behavioral data.

In our case, cross-validation was automatically performed by splitting the annotated video dataset into a training set and a validation set: 80/20 split for Training Sessions (TS) 1–6, and 90/10 for TS7. The validation set allows us to evaluate the model's performance on data it hadn't seen during training, providing an early estimate of generalizability. To evaluate performance, we used the F1 score (but also the accuracy and the confusion matrix).

A binary confusion matrix for each behavior, comparing frame-by-frame predictions to ground truth annotations was also produced by Deg pipeline. This binary confusion matrix shows the number of

True Positives (x=1;y=1) –correctly identified behavior frames, False Positives –behavior predicted but not present (x=1;y=0), True Negatives –correctly identified non-behavior frames (x=0;y=0) and False Negatives –behavior present but missed for each behavior(x=0;y=1).

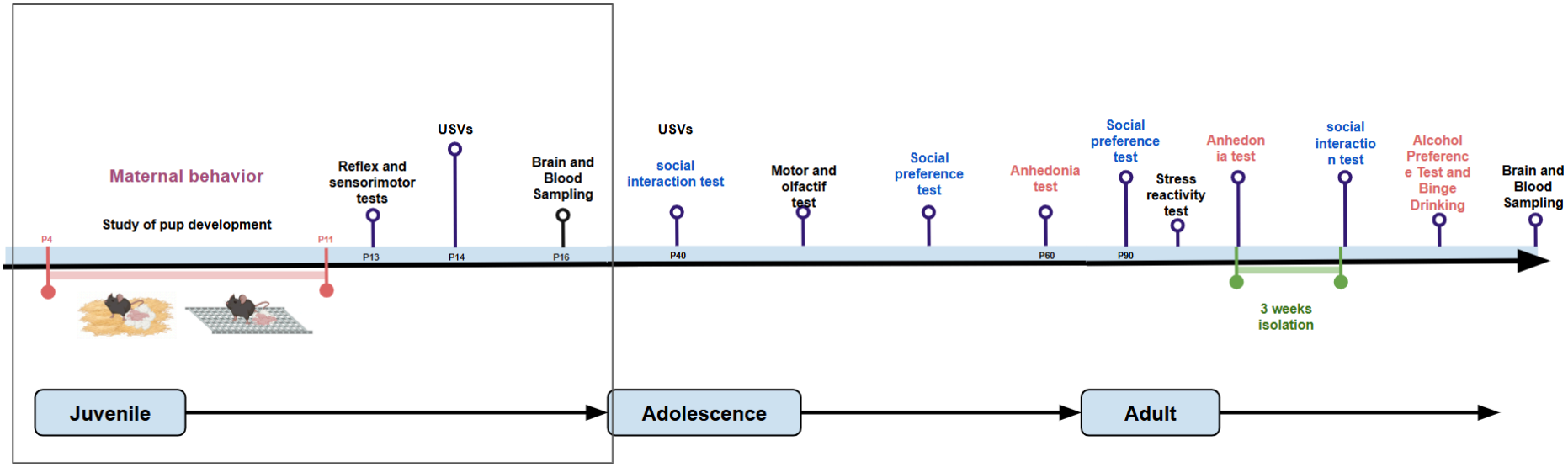


Figure 7. General planning. Process followed by the mice. Sampling needs the mice's scarification, then not all mice follow the same path. It's only a general representation.

LBN pup's available data

To assess early outcomes of stress exposure, previous work in the lab evaluated ultrasonic vocalizations (USVs), motor development, physical maturation, and physiological markers. USVs were recorded at P3 and P14 to assess emotional reactivity and social communication, measuring call number, frequency, duration, and latency. Sensorimotor development was primarily assessed at P12, with daily righting reflex ("Flip") testing from P6–P13; an area under the curve (AUC) from P7–P10 summarized progression. Additional tasks included cliff avoidance and negative geotaxis. Muscle strength and coordination were tested via grasp latency, normalized for weight, and front-limb suspension. Physical maturation milestones included ear twitching, ear and eye opening, walking onset, and body weight (P3, P13). Finally, plasma corticosterone (C0, C1) and mRNA expression of stress- and immune-related genes (e.g., FKBP5, NR3C1, SGK1, SKA2) were quantified to investigate biological correlates of early adversity.

Statistics

Pup data were analyzed by considering three factors: experimental condition (e.g., LBN vs control), cage and then by extent dam, and sex. Cage is the term that will be used further for the dam. The model selection procedure followed a predefined strategy: if the sample size, number of tested pups, was greater than 60 or if Shapiro results on a p-value greater than 0.05 (to ensure sufficient power for two-way comparisons), parametric analysis of variance (ANOVA) models were applied, testing all three factors for each available variable. If the sample size was smaller ($n < 60$) and normality was not followed, the non-parametric Kruskal–Wallis test was used instead. Given the large number of variables and models tested, different Bonferroni corrections were applied to p-values, following the approach described by Bland & Altman : (1) a correction for the number of factors tested per model, (2) a correction for the number of models per variable, and (3) a global correction for the total number of statistical tests performed (Bland et Altman 1995). This multilevel correction strategy allowed for distinguishing variable categories based on whether they passed the same threshold imposed to more lenient or more stringent corrections.

For ANOVA models, results were retained only if the residuals passed a Shapiro–Wilk normality test with a significance threshold of $p < 0.01$, ensuring valid parametric assumptions.

To explore multivariate patterns in pup data, Principal Component Analysis (PCA) was also performed, using the dam (cage) as a grouping factor for visualization purposes, following standard procedures (Jolliffe et Cadima 2016). PCA was applied separately for two sets of variables: (1) variables meeting ANOVA assumptions (more likely to be suitable for linear decomposition and visualization), and (2) a broader set including both ANOVA- and Kruskal–Wallis-compatible

variables, used for unsupervised clustering. Clustering was performed using hierarchical agglomerative clustering (HAC) via the *hclust* function from the R base stats package, following Euclidean distance. Same PCA analysis were also realised with Random forest for confirmation.

For dam behavior data, a combination of Student's t-test, Wilcoxon rank-sum test, ANOVA, and Kruskal–Wallis test was used depending on the distribution and sample size of each variable. Additionally, Bayesian change point analysis was applied to detect significant temporal changes in dam behaviors over time, using the changepoint package in R (Killick et Eckley 2014).

Pup's missing data completion

Early cohorts were exploratory in nature, leading to the progressive inclusion of additional tests in the following cohorts. Consequently, several assessments were not performed in the initial cohorts, resulting in missing data when grouping the cohorts' results. These gaps in the dataset limit the applicability of advanced multivariate analyses, such as principal component analysis (PCA), k-means clustering, or Markov chain modeling. To enable such analyses, it was therefore decided to impute the missing values using robust and comprehensive methods.

We selected the MissForest algorithm, a non-parametric method for imputing missing values. For each variable with missing entries, the algorithm uses the available (non-missing) values as a target and leverages all other variables as predictors in a random forest model. This model then estimates the missing values, repeating the process iteratively across all variables with missing data (Stekhoven et Bühlmann 2012).

Although imputation methods such as MissForest provide powerful solutions for handling missing data, they can be problematic when investigating individual differences, as imputations may obscure true variability or introduce artificial patterns. In the context of this study, these methods were applied primarily to explore the methodological feasibility –to test how well dimensionality reduction or clustering techniques could perform– rather than to draw definitive conclusions or interpretations. Therefore, any inference drawn from these analyses should be considered preliminary, and future studies with complete datasets across all individuals and variables will be necessary to validate these findings and accurately capture biologically meaningful differences.

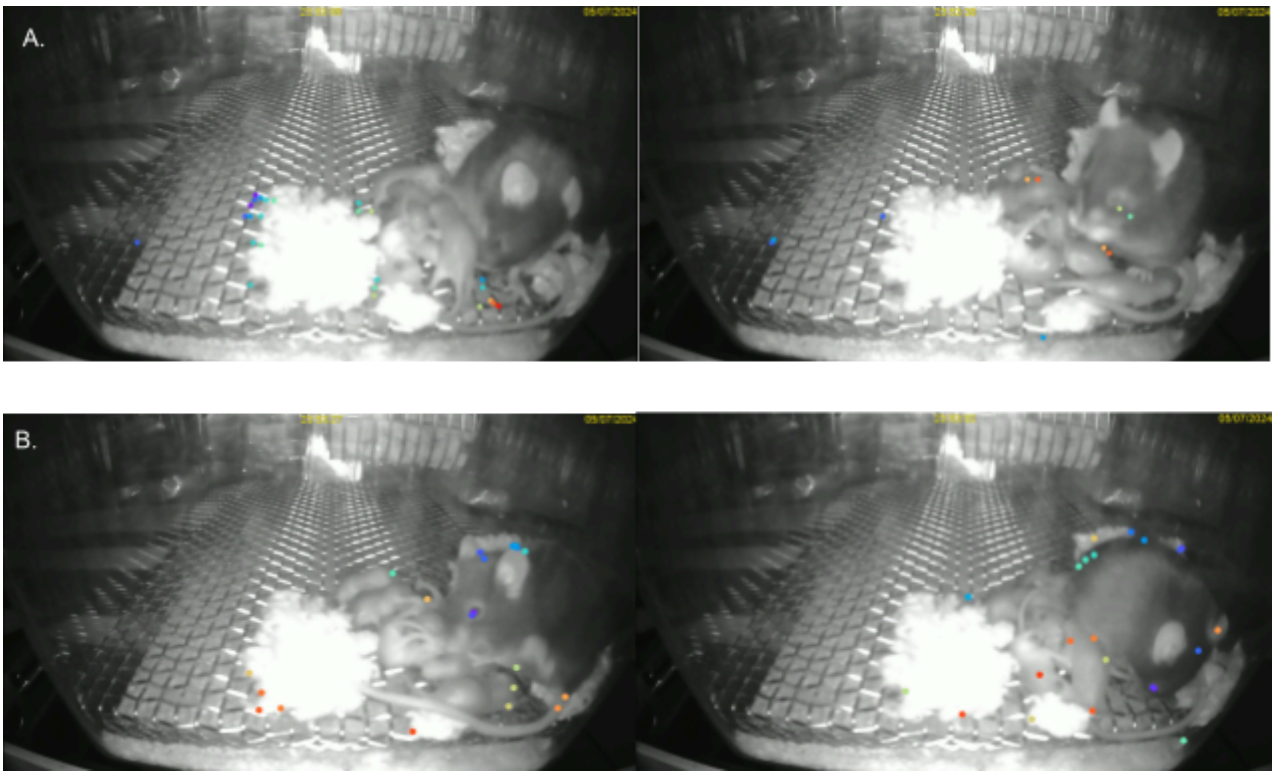


Figure 8. AMBER pose-estimation results. AMBER DeepLabCut (DLC) provides two distinct outputs: one for pup pose-estimation videos (A) and another for dam pose-estimation videos (B). Panel A displays two frames illustrating the output for pups, where each point corresponds to a predicted body part, color-coded by identity, with a total of 9 body parts per pup. Panel B presents the same type of output for dams, with pose estimation across 32 distinct body parts.

Results and discussion

Maternal behaviour automatic detection

AMBER

Fig. 8.A illustrates poor pose-estimation performance for pups. Pose points are present on the nest, or on the dam as are present on the pups. No individual pup has more than two identified keypoints, although the model is expected to detect nine per pup. Similarly, Fig. 8.B shows unreliable pose-estimation for the dam: several predicted key points fall outside the actual body and often correspond to incorrect anatomical locations.

These inaccuracies persist throughout all videos and appear even worse when this analysis is done frame-by-frame. Body points frequently jump between the dam, pups, nest, and even the cage, with no spatial consistency. As suspected, these results confirmed that rat-trained DLC models, including those used by AMBER and on maternal behaviour, are not well-suited for mouse pose-estimation in our experimental conditions.

At this point, two options were considered: (1) improve the DLC models by manually labeling at least 4,500 frames (as done for the AMBER pipeline) and then retrain SimBA behaviour classifiers used by AMBER with a dataset exceeding 3 million frames; or (2) switch to a different behaviour classification approach. Due to time constraints and the availability of direct ethological annotations, we opted for **DeepEthogram** (Deg), a pipeline that bypasses pose estimation entirely and learns behaviour patterns directly from annotated behavioural sequences.

Deg

Deg training was time consuming and is still ongoing, but first training results can be analysed and resulting models can be used carefully and with hindsight.

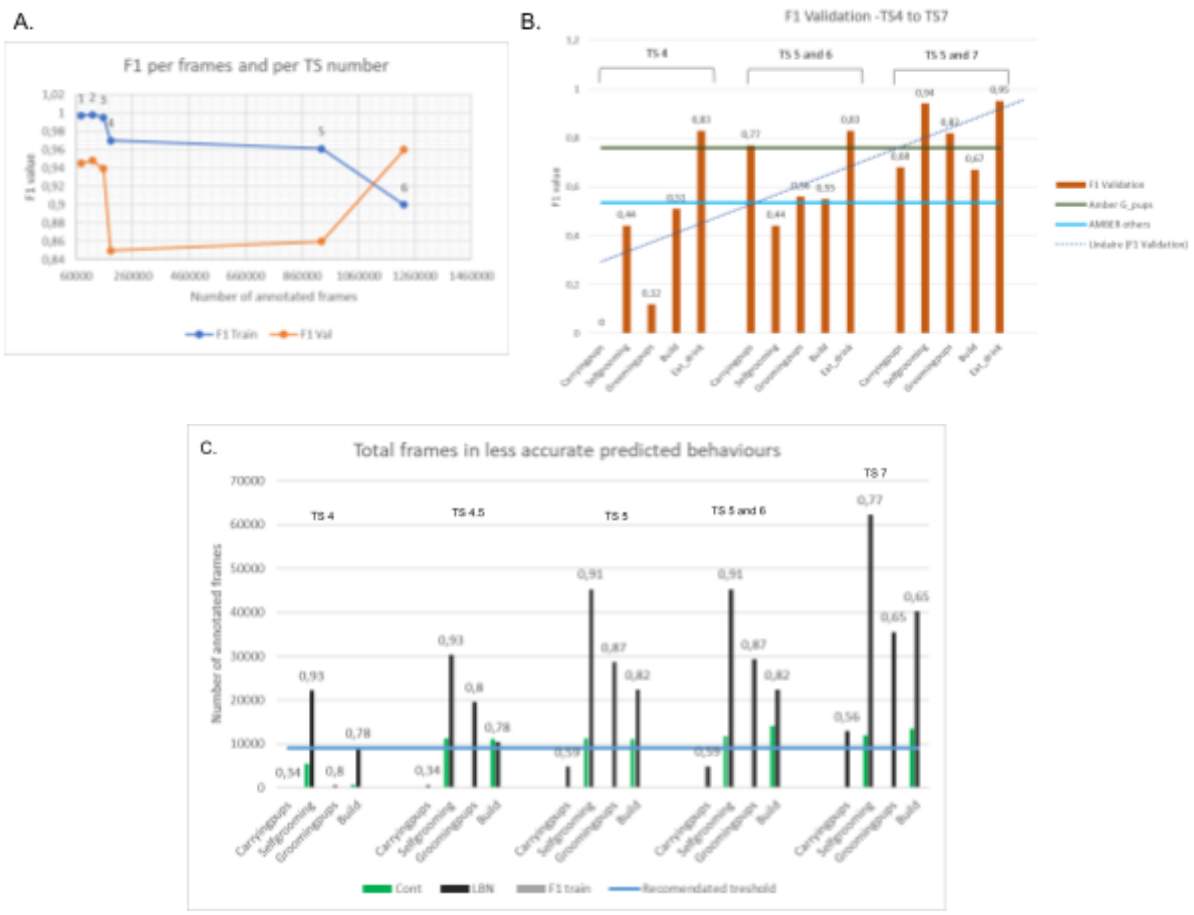


Figure 9. F1 results at different frames present during training and validation. (A) Shows the progression of overall F1 scores for training and validation over sessions. (B) A fitted linear model (dotted line) illustrates this evolution; AMBER's F1 benchmarks appear as solid horizontal lines for two behavior groups. (C) Focuses on four rare behaviors, showing frames and F1 across sessions. "TS 4.5" indicates results before data augmentation.

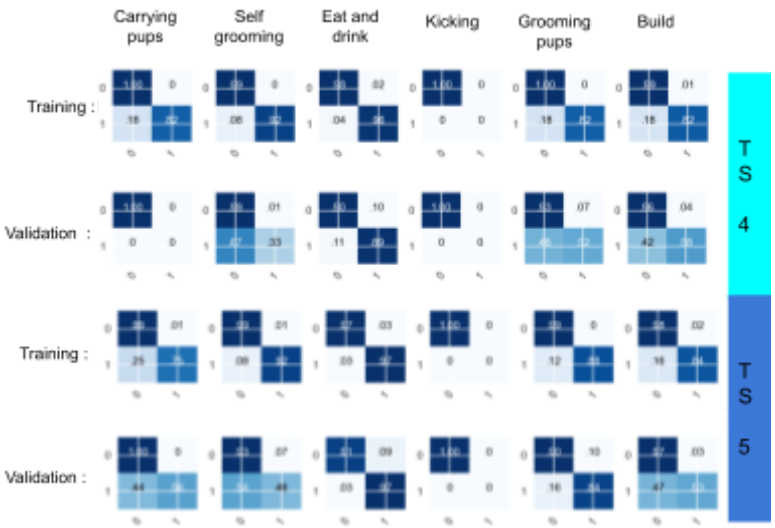


Figure 10. Binary confusion matrix of under presented behaviours between TS. Normalized values from both training and validation are shown. For each behavior, the matrix includes: True Positives (x=1; y=1), False Positives (x=1; y=0), True Negatives (x=0; y=0), and False Negatives (x=0; y=1).

The exploration training, Fig. 9.A, showed us that Deg had **good capabilities for learning to automatically classify behaviors** that have been shown at big rates during training, and that this capability can be ameliorated by increasing the number of frames per behaviour during training. The drop in validation between TS3 and TS 4 performance indicates the point at which rare behaviours– absent from earlier sessions –began to significantly reduce the overall F1 (metric combining precision and recall parameters to assess classification accuracy).

Once we increased the number of frames for rare behaviours, we improved both the **validation overall F1** (0.85 to 0.96), but also the rare behaviours mean F1 (Figure 9.B). As observed in Fig. 9.C, reaching the recommended threshold for frame counts per behaviour allowed us to achieve F1 scores in the **0.75–0.80** range (TS 5 and 6), considered more than acceptable for these cases.

To confirm the superiority of DeG over AMBER in classifying complex behaviours, we show direct comparisons of F1 values for each of them in Fig. 9.C. The model **outperformed AMBER for some behaviours** such as Eating and Drinking, but also achieved good F1 scores for behaviours that AMBER could not detect (e.g., Build, F1 = 0.67). However, some behaviours remain under the recommended threshold (e.g., Carrying pups, < 9,000 real frames) and still underperform compared to AMBER. Carrying pups is particularly problematic, both in terms of training data availability and classification accuracy.

Overall, we observe a **linear improvement in validation F1 across training sessions for most behaviours**, with exceptions like Grooming pups and Self-grooming. However, the **binary confusion matrices (Figure 10)** show clear improvement between TS4 and TS5 for these behaviours. For Grooming pups, the detection rate increased from **52% to 84%**, despite a rise in false positives (from 7% to 10%), while false negatives dropped to **16%**. Similarly, for Self-grooming, false positives increased from **1% to 7%**, while undetected frames were significantly reduced.

Results demonstrate Deg capability of learning complex maternal behaviors and interactions with nest and pups. For our last training, **F1 scores were at 0.90 for training and 0.96 for validation**. Behaviours like *Nest building* or *eating and drinking*, well represented in the training set, tend to reach acceptable performance levels (>0.65). However, Deg models **further training** will be necessary to achieve this level of performance across all behaviours.

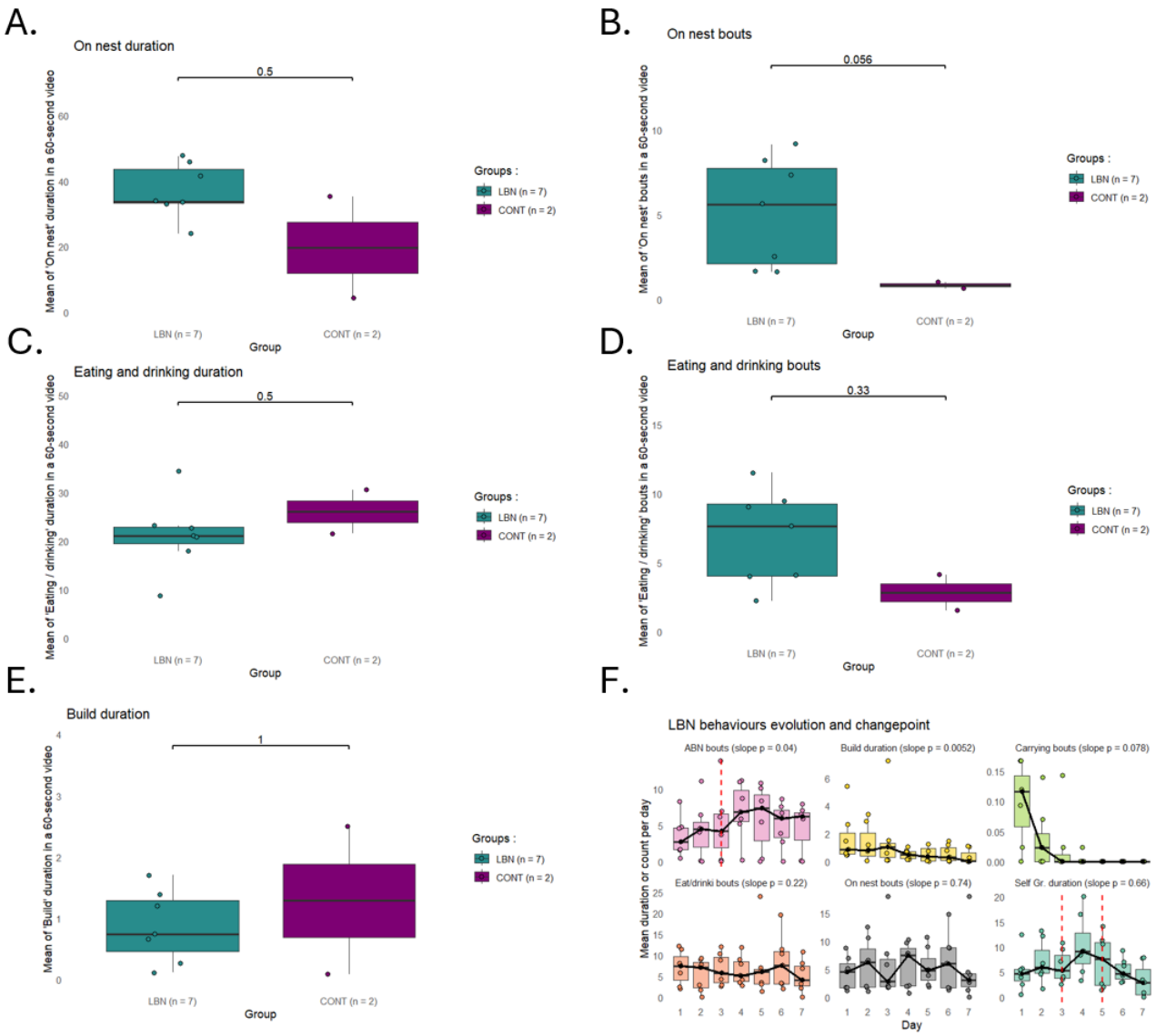


Figure 11. Maternal behaviour durations under LBN and control conditions. (A) Nest attendance duration ratio. (B) On-nest bouts per nest entry. (C) Eating and drinking durations. (D) Eating and drinking counts. (E) Mean nest-building durations. (F) Daily evolution of selected maternal behaviours under LBN. A changepoint (Killick et Eckley 2014) day was identified using Pruned Exact Linear Time (PELT).

LBN animal model validation

Maternal results

Analysis of expressed behaviours in seven LBN cages compared to n two control cages show a tendency of modified maternal behaviours by LBN conditions. The analysis of more cages is ongoing using the last trained version of Deg.

As shown in Figure 11, although non-parametric tests do not yield statistically significant differences due to the low number of observations collected until now, some behavioral trends emerge between LBN and control groups. LBN dams tend to spend more time on the nest, with a median of 35 seconds per video, corresponding to approximately 55% of the observed time (Figure 11.A). In contrast, control dams show a lower median of 22 seconds.

Interestingly, the increase in On Nest duration observed in LBN dams is not mirrored by a reduction in the number of On nest bouts. Since each bout represents a transition, an increase of total On Nest duration between groups would normally imply fewer transitions in LBN dams. However, despite spending more time on the nest, LBN dams exhibit a higher number of transitions compared to controls. Indicating that their behavior is more fragmented (Figure 11.B).

Eating and drinking behaviors are broadly similar across conditions, though LBN dams display a slightly higher number of bouts which may reflect more fragmented patterns of behavior (Figure 11.C,D). Nest-building behavior is observed in both groups but appears more pronounced in control dams, likely due to greater availability of nesting material.

Daily trends in maternal behaviors under LBN conditions remain relatively stable. A linear decrease is observed exclusively for Building, while arched-back nursing (ABN)—shows a linear increase (Figure 11.F). PELT analysis detected a changepoint on LBN day 3 for both ABN bouts and self-grooming durations. For ABN, this point marks an increase followed by behavioral stabilization. Self-grooming increases after LBN day 3 potentially reflecting heightened stress or anxiety in LBN dams. These elevated levels persist for approximately two days (LBN day 4–5), before returning to baseline at LBN day 6.

Consequences on pups

ANOVA and kruskal-wallis statistical tests of pup behavioural, transcriptomic, physical, motor, and neurochemical data reveal distinct phenotypic characteristics associated with LBN exposure.

Variable	Modele	Facteur	P_value	P_adj	Test	Shapiro_p	P_adj_model	P_adj_global
<chr>	<chr>	<chr>	<dbl>	<dbl>	<chr>	<dbl>	<dbl>	<dbl>
flip.AUC.p7.10	Modèle_1	Condition	6.574333e-07	6.771563e-05	Kruskal-Wallis	NA	6.574333e-07	1.972300e-06
norm.av.front.lat	Modèle_1	Condition	5.541065e-04	5.707297e-02	Kruskal-Wallis	NA	5.541065e-04	1.662320e-03
d.ear.c.open	Modèle_1	Condition	1.991389e-03	2.051131e-01	Kruskal-Wallis	NA	1.991389e-03	3.982778e-03
d.eye.l.open	Modèle_1	Condition	4.349140e-04	4.479614e-02	Kruskal-Wallis	NA	4.349140e-04	1.304742e-03
d.transition	Modèle_1	Condition	2.876359e-03	2.962649e-01	Kruskal-Wallis	NA	2.876359e-03	8.629076e-03
d.walk	Modèle_1	Condition	1.711485e-10	1.762830e-08	Kruskal-Wallis	NA	1.711485e-10	5.134455e-10
tot.n.voc.p14	Modèle_1	Condition	5.692381e-05	5.863153e-03	Kruskal-Wallis	NA	5.692381e-05	1.707714e-04
av.ep.dur.p14	Modèle_1	Condition	7.172813e-03	7.387997e-01	Kruskal-Wallis	NA	7.172813e-03	2.151844e-02
lat.p14	Modèle_1	Condition	6.390475e-03	6.582190e-01	Kruskal-Wallis	NA	6.390475e-03	1.917143e-02
Cortp	Modèle_1	Condition	9.873332e-03	1.000000e+00	Kruskal-Wallis	NA	9.873332e-03	2.962000e-02
FKBP5	Modèle_1	Condition	2.738684e-03	2.820845e-01	ANOVA	0.658978	2.738684e-03	1.095474e-02
slope	Modèle_1	Condition	3.063343e-03	3.155243e-01	Kruskal-Wallis	NA	3.063343e-03	9.190029e-03

Figure 12. Multifactorial ANOVA results for pup variables affected by LBN. Four models were tested per variable: (1) Condition, (2) Cage ID/Dam, (3) Condition × Cage ID/Dam, and (4) Condition × Cage ID/Dam × Sex. Table sorted by adjusted P-value. Bonferroni corrections: per model (number of factors), per variable (across models), and for all variables (all tests across all variables).

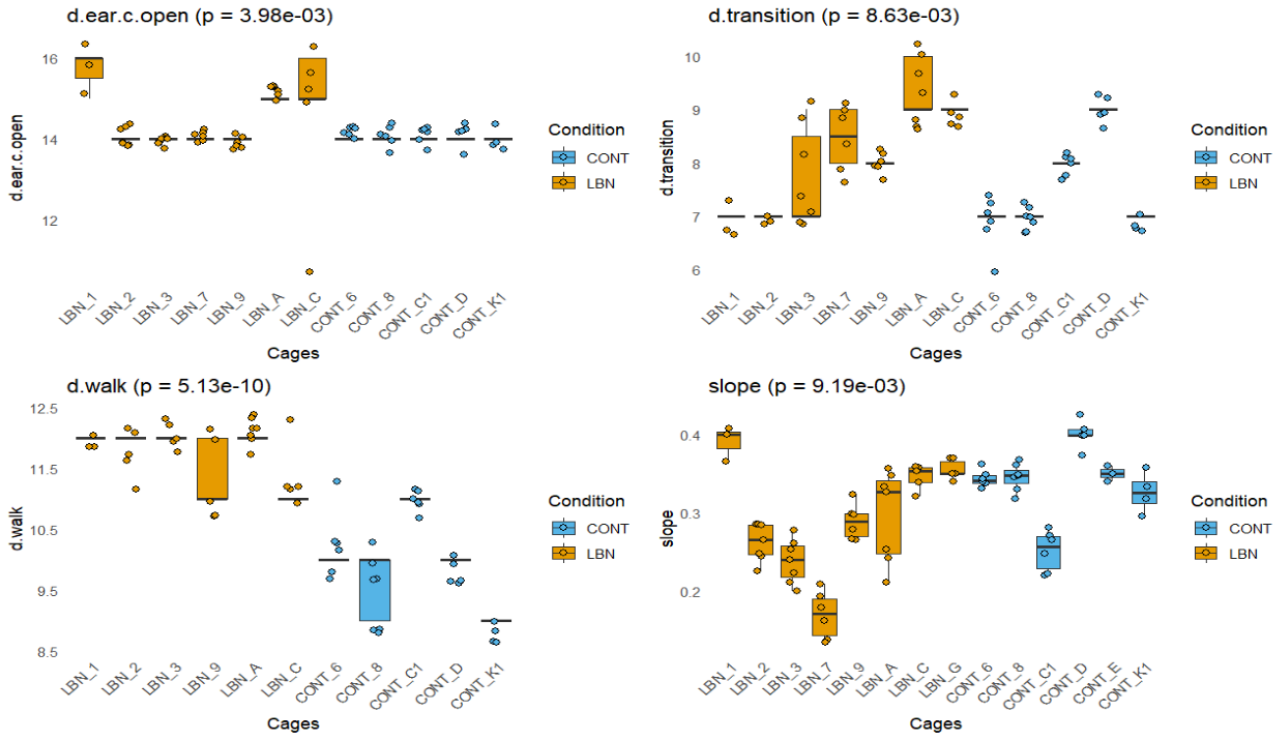


Figure 13. Box plot of 4 most significant variables affected specifically by control or LBN dams. For each variable inter-group (LBN/CONT) adjusted p-value are reported in the graph title (NS is not significant, then p-value > 0.05). A cage corresponds to one dam, and then to specific maternal cares.

Indeed, Fig. 12 highlights 12 over 31 variables found to differ significantly between LBN and control groups. These include, among others, one variable related to transcriptomic profiles, five concerning motor development, and 3 pertaining to physical characteristics. Although no significant differences were observed in plasma corticosterone levels between groups, expression changes in genes involved in the hypothalamic–pituitary–adrenal (HPA) axis suggest molecular alterations. Among these, FKBP5 expression was slightly reduced in LBN-exposed pups (adjusted global p-value = 0.04). In terms of motor development, the onset of walking was significantly delayed under LBN conditions, with a mean delay of nearly two days compared to controls ($p < 0.001$). Physical development was also impacted by LBN, indeed bodyweight slope was significantly reduced for LBN mice ($p < 0.05$).

Maternal specific behaviours have specific consequences on pups phenotypes

LBN pups are different depending on maternal care

Statistical tests showed that **dam identity significantly modulates pup phenotypes**, independently of overall condition (LBN or Control). **12 variables were found** (not the same as previous 12) as significant for this modulation.

The **four most dam-impacted variables** (All variables adjusted p-value < 0.05) are illustrated in figure 13. Intra-LBN variability is observed for the *first day of ear opening,* for *bodyweight slope*, and for the *first day of walking* and *first day of transition*. Inter-group control/LBN variability is also significant for each of these variables. For example, for the bodyweight slope variable we observe that LBN dam 7 shows a very low slope compared to other LBN cages, same for cage number 3. For the ear opening variable, the delay exists only for LBN cages A, C and 1 (no data for cage 7 and 3).

To explore the structure of this phenotypic variability in a more global way, we performed a **Principal Component Analysis (PCA)** for the two categories : **condition-sensitive variables** (where significant differences were observed between LBN and control) and **dam-sensitive variables** shown on figure 12 (where significant differences were observed between cages for an specific condition).

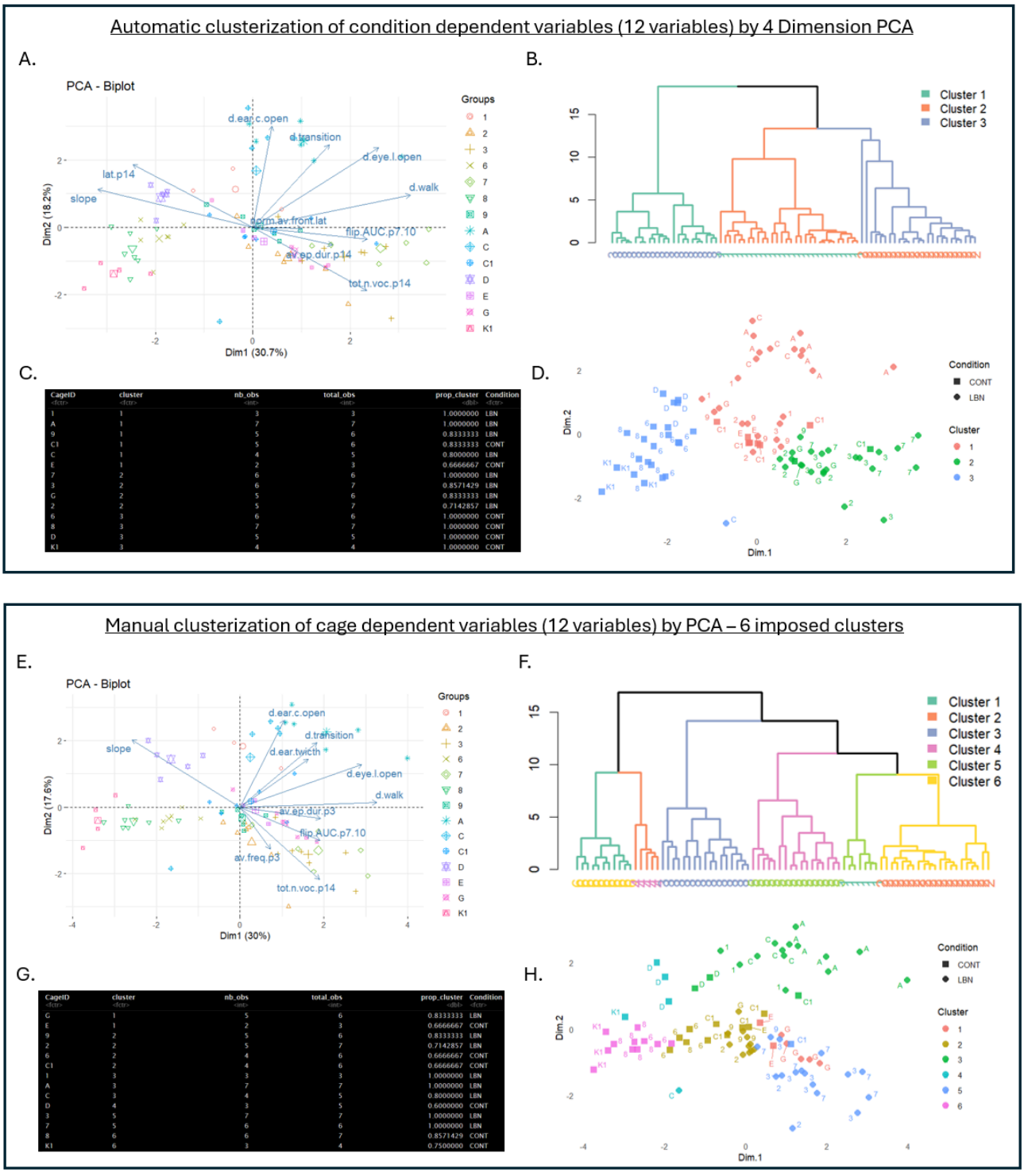


Figure 14. PCA and hierarchical clustering results based on two groups of variables : condition-dependent (A–D) and dam-dependent (E–H). (A, E) Two-dimensional PCA biplots showing both variable vectors and individual coordinates. (B, F) Dendrograms from hierarchical clustering using the first 4 principal components. (C, G) Tables indicating, for each cage, the assigned cluster and the proportion of individuals belonging to it. (D, H) PCA plots with individuals colored by cluster and shaped by condition or cage identity. Cluster assignment was performed using either automatic dynamic tree cutting (A–D) or fixed-height cut tree (E–H).

Condition and dam-dependent variables exhibit distinct distributions in PCA space (Figure 14.A, E). Cluster assignments from dynamic tree cutting (**condition variables**) and imposed clusters tree cutting (**dam variables**) were coherent, consistently **segregating LBN and Control cages** (Figure 14.C, G). Most combinations produced two clusters (one LBN, one Control), but **condition-dependent** variables revealed **three clusters** via both PCA-based (Figure 14.C, D) and random forest clustering (Not presented in these results). Resulted *Cluster 1* contains four LBN and two Control cages; *Cluster 2* includes the remaining LBN cages; and *Cluster 3* comprises the rest of the Control cages.

Projected onto PCA dimensions (Figure 14.D), Cluster 1 individuals load weakly on PC1; characterized by high vocalizations, reduced forelimb strength, and delayed walking; indicating a milder phenotype. Cluster 2 (exclusively LBN) shows stronger expression of these traits, reflecting more pronounced developmental delays and stress behaviors.

Intra group (LBN or control) clustering further highlighted patterns are: LBN cages 7 and 3 consistently grouped in Cluster 5 (Figure 14.G, H). Cages A, C, and 1 formed Cluster 3, indicating co-specificity. Other LBN cages overlapped partially with both groups, suggesting intermediate profiles. Overall, pup phenotypes under LBN are heterogeneous and **cluster by individual maternal care**. Specifically, pups from dams **3 and 7 show more severe phenotype**s : reduced weight gain (Figure 14.E) and increased USVs.

Discussion

In this study, we developed and validated a **novel pipeline for the automatic annotation of continuous home-cage videos capturing dam–pup interactions under two experimental conditions.** By leveraging DeepEthogram with our custom-trained models, we achieved frame-by-frame classification of both individual and complex interaction behaviours. Notably, our latest model reached performance levels comparable to well-established rat-oriented methods such as AMBER, despite relying on only 30 % of their annotated training frames. Because our approach does not depend on explicit animal body-part localization, we successfully detected an additional category of dam–nest interactions (nest building) that had previously been inaccessible.

During validation, core maternal behaviours were robustly identified. Indeed, our models detected must behaviors with an F1 superior to **0.95.** Although for some behaviours the F1 score was lower, the threshold of **0.65** is surpassed by all of them. Nevertheless, it'll be important to augmentate low represented behaviours videos to increase the F1 and ameliorate their detection. With the developed method, we've characterised maternal behaviors of 12 cages, 7 on LBN conditions and 5 on normal ones. The analysis of these results confirmed that LBN condition affects maternal

behaviors by inducing fragmented maternal care resulting in higher On nest transitions, elevated eat and drink bouts, but also increasing anxiety-like behaviours.

In parallel, **pup outcomes are also modified by LBN conditions**. Different aspects such as body weight, genetic expression of stress-implicated proteins, developmental aspects, and motor and strength performances are affected. These modifications are generally significant, but individual pups vary: some are more affected in certain measures, and others in different measures. These differences likely reflect inter-individual variability; however, the results suggest that the identity of the dam with which pups were raised plays a major role in shaping certain outcomes. Notably, pups from cages 3 and 7 exhibited similar phenotypic profiles. Interestingly, dam 3 was the only one to express kicking behaviour, and further analyses revealed strong behavioural similarities between dams 3 and 7, pointing toward a potential shared caregiving pattern that may underlie the observed pup phenotypes.

Conclusions and perspectives

Taken together, our results establish that automated behavioural classification in the home-cage context can robustly quantify maternal-care patterns that could be predictive of a broad spectrum of pup developmental outcomes.

Building upon these findings, we envisage several avenues for further investigation. First, longitudinal studies should extend beyond the early postnatal period to follow pup cohorts into **adolescence and adulthood**. Such work would clarify how early maternal fragmentation modulates later phenotypes. Second, integrating our Deg pipeline with machine-learning-based predictive models (e.g., random forest regressors, multilayer perceptrons with hidden Markov chains and Vector AutoRegression) could enable early forecasting of **individual pup trajectories based on maternal-care metrics**. Third, deploying this technology within more naturalistic settings such as enriched home cages as EcoHab would permit observation of adolescent and adult social dynamics, approximating ethological conditions more closely. By combining automated tracking with robust behavioural classifiers (Deg, DLC-SimBA), we can elucidate how **complex social networks and environmental affordances modulate developmental outcomes**.

In conclusion, our study underscores the feasibility and utility of high-throughput, automated analysis of dam–pup interactions. By enabling precise, scalable quantification of maternal behaviour, this approach opens the door to mechanistic studies on how early life caregiving patterns shape neurodevelopmental trajectories. In the long term, such methods may inform translational efforts to identify **biomarkers of vulnerability and resilience in at-risk populations**, ultimately guiding interventions that can mitigate the enduring effects of early neglect.

Bibliographie

Ayre, Kay, et Govind Krishnamoorthy. 2020. *Trauma Informed Behaviour Support: A Practical Guide to Developing Resilient Learners*. University of Southern Queensland. https://doi.org/10.26192/q5zy0.

Bath, K., G. Manzano-Nieves, et H. Goodwill. 2016. « Early Life Stress Accelerates Behavioral and Neural Maturation of the Hippocampus in Male Mice ». *Hormones and Behavior* 82 (juin):64‑71. https://doi.org/10.1016/j.yhbeh.2016.04.010.

Bland, J. Martin, et Douglas G. Altman. 1995. « Multiple Significance Tests: The Bonferroni Method », janvier. https://doi.org/10.1136/bmj.310.6973.170.

Bohnslav, James P, Nivanthika K Wimalasena, Kelsey J Clausing, Yu Y Dai, David A Yarmolinsky, Tomás Cruz, Adam D Kashlan, et al. 2021. « DeepEthogram, a machine learning pipeline for supervised behavior classification from raw pixels ». *eLife* 10:e63377. https://doi.org/10.7554/eLife.63377.

Bryda, Elizabeth C. 2013. « The Mighty Mouse: The Impact of Rodents on Advances in Biomedical Research ». *Missouri Medicine* 110 (3): 207‑11.

Demaestri, Camila, Meghan Gallo, Elisa Mazenod, Alexander T. Hong, Hina Arora, Annabel K. Short, Hal Stern, Tallie Z. Baram, et Kevin G. Bath. 2022. « Resource scarcity but not maternal separation provokes unpredictable maternal care sequences in mice and both upregulate Crh-associated gene expression in the amygdala ». *Neurobiology of Stress* 20 (septembre):100484. https://doi.org/10.1016/j.ynstr.2022.100484.

Duffy, Korrina A., Katie A. McLaughlin, et Paige A. Green. 2018. « Early life adversity and health-risk behaviors: proposed psychological and neural mechanisms ». *Annals of the New York Academy of Sciences* 1428 (1): 151‑69. https://doi.org/10.1111/nyas.13928.

Gallo, Meghan, Daniel G. Shleifer, Livea D. Godoy, Dayshalis Ofray, Aliyah Olaniyan, Talia Campbell, et Kevin G. Bath. 2019. « Limited Bedding and Nesting Induces Maternal Behavior Resembling Both Hypervigilance and Abuse ». *Frontiers in Behavioral Neuroscience* 13 (juillet). https://doi.org/10.3389/fnbeh.2019.00167.

Haller, J., G. Harold, C. Sandi, et I. D. Neumann. 2014. « Effects of Adverse Early-Life Events on Aggression and Anti-Social Behaviours in Animals and Humans ». *Journal of Neuroendocrinology* 26 (10): 724‑38. https://doi.org/10.1111/jne.12182.

Herman, James P. 2022. « The Neuroendocrinology of Stress: Glucocorticoid Signaling Mechanisms ». *Psychoneuroendocrinology* 137 (mars):105641. https://doi.org/10.1016/j.psyneuen.2021.105641.

Hosseinichimeh, Niyousha, Hazhir Rahmandad, et Andrea K. Wittenborn. 2015. « Modeling the hypothalamus-pituitary-adrenal axis: A review and extension ». *Mathematical biosciences* 268 (octobre):52‑65. https://doi.org/10.1016/j.mbs.2015.08.004.

Jolliffe, Ian T., et Jorge Cadima. 2016. « Principal component analysis: a review and recent developments ». *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences* 374 (2065): 20150202. https://doi.org/10.1098/rsta.2015.0202.

Killick, Rebecca, et Idris A. Eckley. 2014. « Changepoint: An R Package for Changepoint Analysis ». *Journal of Statistical Software* 58 (juin):1‑19. https://doi.org/10.18637/jss.v058.i03.

König, Barbara, et Hubert Markl. 1987. « Maternal Care in House Mice ». *Behavioral Ecology and Sociobiology* 20 (1): 1‑9. https://doi.org/10.1007/BF00292161.

Lapp, Hannah E., Melissa G. Salazar, et Frances A. Champagne. 2023. « Automated Maternal Behavior during Early Life in Rodents (AMBER) Pipeline ». *Scientific Reports* 13 (1): 18277. https://doi.org/10.1038/s41598-023-45495-4.

Lautarescu, Alexandra, Michael C. Craig, et Vivette Glover. 2020. « Prenatal Stress: Effects on Fetal and Child Brain Development ». *International Review of Neurobiology* 150:17‑40.

https://doi.org/10.1016/bs.irn.2019.11.002.

Lee, Seung Hyun, et Eui-Man Jung. 2024. « Adverse Effects of Early-Life Stress: Focus on the Rodent Neuroendocrine System ». *Neural Regeneration Research* 19 (2): 336. https://doi.org/10.4103/1673-5374.377587.

Lynch, Carol Becker, et Bernard P. Possidente. 1978. « Relationships of maternal nesting to thermoregulatory nesting in house mice (*Mus musculus*) at warm and cold temperatures ». *Animal Behaviour* 26 (novembre):1136‑43. https://doi.org/10.1016/0003-3472(78)90103-3.

Papadimitriou, Anastasios, et Kostas N. Priftis. 2009. « Regulation of the Hypothalamic-Pituitary-Adrenal Axis ». *Neuroimmunomodulation* 16 (5): 265‑71. https://doi.org/10.1159/000216184.

Shorten, Connor, et Taghi M. Khoshgoftaar. 2019. « A survey on Image Data Augmentation for Deep Learning ». *Journal of Big Data* 6 (1): 60. https://doi.org/10.1186/s40537-019-0197-0.

Stekhoven, Daniel J., et Peter Bühlmann. 2012. « MissForest—Non-Parametric Missing Value Imputation for Mixed-Type Data ». *Bioinformatics* 28 (1): 112‑18. https://doi.org/10.1093/bioinformatics/btr597.

Sturman, Oliver, Lukas von Ziegler, Christa Schläppi, Furkan Akyol, Mattia Privitera, Daria Slominski, Christina Grimm, et al. 2020. « Deep learning-based behavioral analysis reaches human accuracy and is capable of outperforming commercial solutions ». *Neuropsychopharmacology* 45 (11): 1942‑52. https://doi.org/10.1038/s41386-020-0776-y.

Veenema, Alexa H. 2009. « Early life stress, the development of aggression and neuroendocrine and neurobiological correlates: What can we learn from animal models? » *Frontiers in Neuroendocrinology*, Hormones & Social Behavior, 30 (4): 497‑518. https://doi.org/10.1016/j.yfrne.2009.03.003.

Walker, Claire-Dominique, Kevin G. Bath, Marian Joels, Aniko Korosi, Muriel Larauche, Paul J. Lucassen, Margaret J. Morris, et al. 2017. « Chronic early life stress induced by limited bedding and nesting (LBN) material in rodents: critical considerations of methodology, outcomes and translational potential ». *Stress (Amsterdam, Netherlands)* 20 (5): 421‑48. https://doi.org/10.1080/10253890.2017.1343296.

Weber, Elin M., et I. Anna S. Olsson. 2008. « Maternal behaviour in *Mus musculus* sp.: An ethological review ». *Applied Animal Behaviour Science* 114 (1): 1‑22. https://doi.org/10.1016/j.applanim.2008.06.006.

WHO, et Malin Bring. 2015. « Violence and Injuries EURO ». 2015. https://www.who.int/europe/health-topics/violence/preventing-child-maltreatment.

WHO, et S.Becker. 2024. « Child Maltreatment ». 2024. https://www.who.int/news-room/fact-sheets/detail/child-maltreatment.

Worley, Nicholas B., Anthony Djerdjaj, et John P. Christianson. 2019. « Convolutional Neural Network Analysis of Social Novelty Preference Using DeepLabCut ». bioRxiv. https://doi.org/10.1101/736983.

Abstract

Maternal neglect is the most common form of child abuse. It affects offspring neurodevelopment and is associated with numerous behavioral and neuroendocrinological alterations, increasing the risk of psychiatric disorders in both childhood and adulthood. However, characterizing the progression of symptoms resulting from early-life adversity remains challenging in humans due to the scarcity of data from the earliest years of maltreatment.

To address this limitation, a well-characterized mouse model was used to induce fragmented maternal care by housing dams in a limited bedding and nesting (LBN) environment, which alters maternal behavior. In this study, machine learning models were developed and trained to identify specific maternal behaviors observed within the LBN paradigm. By integrating data from previous tests, we identified distinct groups of LBN-exposed pups, with differences driven by variations in maternal behavior.

Our models successfully classified behaviors including arched-back nursing, licking and grooming of pups, pup retrieval (carrying pups), self-grooming, feeding and drinking, nest presence (on and off nest), and nest building. We surpassed the performance of well-established methods such as AMBER and were also able to detect previously uncharacterized behaviors with good accuracy, including nest building and pup retrieval ($F1 = 0.67$).

As these methodologies continue to develop, they will be applied to the analysis of adolescent and adult data, as well as new experimental cohorts. Our results highlight a link between maternal behaviors and pup outcomes, opening the door not only to predictive modeling of offspring trajectories but also to the continuous home-cage monitoring and detection of behaviors associated with resilience or vulnerability in adult mice exposed to LBN conditions.