



## OPEN ACCESS

EDITED BY  
Stefano Gaburro,  
Tecniplast, Italy

REVIEWED BY  
Damien Huzard,  
Université de Montpellier, France  
Eoin C. O'Connor,  
Roche Innovation Center, Switzerland

\*CORRESPONDENCE  
Jun-Il Yoo  
✉ furim@daum.net

RECEIVED 10 July 2025  
ACCEPTED 05 August 2025  
PUBLISHED 22 August 2025

## CITATION

Bhola S, Kim H-B, Kim HS, Gu B and Yoo J-I  
(2025) Does advancement in marker-less  
pose-estimation mean more quality  
research? A systematic review.  
*Front. Behav. Neurosci.* 19:1663089.  
doi: 10.3389/fnbeh.2025.1663089

## COPYRIGHT

© 2025 Bhola, Kim, Kim, Gu and Yoo. This is  
an open-access article distributed under the  
terms of the [Creative Commons Attribution  
License \(CC BY\)](#). The use, distribution or  
reproduction in other forums is permitted,  
provided the original author(s) and the  
copyright owner(s) are credited and that the  
original publication in this journal is cited, in  
accordance with accepted academic  
practice. No use, distribution or reproduction  
is permitted which does not comply with  
these terms.

# Does advancement in marker-less pose-estimation mean more quality research? A systematic review

Shivam Bhola<sup>1,2</sup>, Hyun-Bin Kim<sup>3</sup>, Hyeon Su Kim<sup>3</sup>, BonSang Gu<sup>3</sup>  
and Jun-Il Yoo<sup>1,2\*</sup>

<sup>1</sup>Department of Orthopedic Surgery, Inha University Hospitals, Incheon, Republic of Korea,

<sup>2</sup>Department of Biomedical Sciences, College of Medicine, and Program in Biomedical Science &  
Engineering, Inha University, Incheon, Republic of Korea, <sup>3</sup>Department of Biomedical Research  
Institute, Inha University Hospitals, Incheon, Republic of Korea

Recent breakthroughs in marker-less pose-estimation have driven a significant transformation in computer-vision approaches. Despite the emergence of state-of-the-art keypoint-detection algorithms, the extent to which these tools are employed and the nature of their application in scientific research has yet to be systematically documented. We systematically reviewed the literature to assess how pose-estimation techniques are currently applied in rodent (rat and mouse) models. Our analysis categorized each study by its primary focus: tool-development, method-focused, and study-focused studies. We mapped emerging trends alongside persistent gaps. We conducted a comprehensive search of Crossref, OpenAlex PubMed, and Scopus for articles published on rodent pose-estimation from 2016 through 2025, retrieving 16,412 entries. Utilizing an AI-assisted screening tool, we subsequently reviewed the top ~1,000 titles and abstracts. 67 papers met our criteria: 30 tool-focused reports, 28 method-focused studies, and nine study-focused papers. Publication frequency trend has accelerated in recent years, with more than half of these studies published after 2021. Through a detailed review of the selected studies, we charted emerging trends and key patterns, from the emergence of new keypoint-detection methods to their integration into behavioral experiments and adoption in various disease contexts. Despite significant progress in marker-less pose-estimation technologies, their widespread application remains limited. Many laboratories still rely on traditional behavioral assays, under-using advanced tools. Establishing standardized protocols is the key step to bridge this gap, which will ultimately realize the full potential of marker-less pose-estimation and even greater insight into preclinical behavioral science.

## KEYWORDS

marker-less pose estimation, keypoint detection, behavior classification, rodent model, systematic review

# 1 Introduction

The fate preclinical behavioral science relies heavily on early *in vivo* experimentations. Detailed quantification of rodent behavior is essential for understanding disease progression, and treatment efficacy (Lauer et al., 2022). Traditionally, researchers have relied on simple, manual assays such as timing a mouse's pause before exploration or counting how often it crosses grid lines that require human observers to note every action. Besides being tedious and prone to bias, these manual approaches usually miss subtle micro-behaviors such as tiny head lifts, brief standing events, or slight changes in stride that can contain critical clues about early pathological signs (Miller et al., 2011; Desland et al., 2014). Now the question is about the way to detect those subtle micro-behaviors.

Over the last ten years, deep-learning-powered, marker-less pose-estimation has transformed behavioral analysis by detecting key anatomical points like the snout, paws, and tail from video footage without any physical markers. The DeepLabCut (DLC) software (2018) achieved human-level accuracy in tracking fast-moving rodents at the pixel scale trained using 50–200 manually labeled frames (Mathis et al., 2018). Following this innovation, a broad list of tools has emerged, including, AlphaTracker (Chen et al., 2023), DeepLabStream (Schweihoff et al., 2021), Keypoint-MoSeq (Weinreb et al., 2024), and Social LEAP Estimates Animal Poses [SLEAP (Pereira et al., 2022)].

Although these pose-estimation tools have advanced rapidly, their adoption in standard rodent research workflows remains sporadic and largely undocumented. Many labs still depend on hand-scored tests, missing out on the detailed, high-resolution data that automated pose-estimation algorithms can offer. Key contributing factors include: (1) operational costs such as retraining personnel, reconfiguring equipment, and provisioning computational resources require significant effort (Hagelskjær et al., 2018); (2) technical complexity such as diverse pose-estimation packages demand different installations and parameter settings, non-expert users can become easily overwhelmed (Dubey and Dixit, 2023); (3) a lack of standards leaving users without clear guidelines for evaluation tools (Reed et al., 1999); and (4) video-data challenges, the burden of archiving large datasets and ensuring analyses can be reliably repeated (Nassauer and Legewie, 2019).

To map the current state of adoption, we systematically reviewed studies published in last 10 years by searching four databases Crossref, OpenAlex, PubMed, and Scopus for reports of marker-less, keypoint-based tracking in rat and mouse models. Using an AI-assisted screening process, we narrowed approximately 16,412 initial hits to 63 relevant articles and four Tool-focused papers manual included. Each was classified into one of three categories: tool-focused, method-focused, or study-focused study and we mapped key trends and remaining challenges.

Our specific aims were to (1) quantify publication trends and the prevalence of different software platforms; (2) classify research by its primary aim and experimental setting [such as locomotion assays (Sturman et al., 2020), social behavior tests (Sterley et al., 2024), and anxiety paradigms (Sharma et al., 2025)]; (3) map the behavioral assays and disease contexts [such as Parkinson's (Andreoli et al., 2021), Alzheimer's (Miller et al., 2024), and pain models (Li et al., 2024)] that have used pose tracking; and (4)

highlight gaps and opportunities. Through this in-depth analysis, we aim to assess whether the accuracy, flexibility, and ease of use of current pose-estimation tools have been broadly adopted in *in vivo* research, as well as the gaps between innovative technologies and standard preclinical behavioral science.

# 2 Materials and method

Following PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) guidelines, we systematically reviewed only publications focused on rodent marker-less pose-estimation techniques. Below, we detail the protocol we followed for the search strategy, eligibility criteria, study selection, data extraction, risk of bias assessment, and data synthesis

## 2.1 Search strategy

We performed a comprehensive literature search to identify studies involving marker-less pose-estimation in rodent (rat or mouse) models. The search covered publications from January 2016 up to March 2025. We constructed search queries using combinations of keywords related to pose-estimation (such as “pose-estimation,” “posture tracking,” “keypoint detection,” “behavioral tracking”) and rodents (such as “rodent,” “rat,” “mouse,” “mice”), as well as specific tool names known in the field (such as “DeepLabCut,” “LEAP,” “SLEAP,” etc.). These searches were run across multiple databases and search engines, including Crossref, OpenAlex PubMed, and Scopus. To facilitate broad retrieval, we utilized the Publish or Perish software to query those databases with standardized search strings (Harzing, 2007). The search was limited to the English language literature. The final search was completed on 18 March 2025. All references retrieved were imported into a reference manager, and duplicate entries were removed prior to screening. The process was completely automated without any human intervention.

## 2.2 Eligibility criteria

Studies identified from the search were evaluated against predetermined inclusion and exclusion criteria:

Inclusion criteria: (1) Studies must involve marker-less pose-estimation (keypoint-based tracking of body parts) performed on rodents (rats or mice). (2) The pose-estimation should be applied to data from an experimental or observational study (i.e., the researchers collected or used rodent behavioral data in which pose tracking was implemented). (3) The pose tracking must involve more than one key point (to exclude cases like single-point tracking of an animal's centroid). (4) Articles must present original data rather than hypothetical concepts.

Exclusion criteria: (1) Review papers, editorials, or meta-analyses were excluded (we only included primary research studies). (2) Studies focusing on non-rodent species or on humans were excluded, even if they discussed pose-estimation, to keep the scope specific to rodent models. (3) Studies that did not actually perform keypoint detection – for example, those that only

discuss pose-estimation conceptually or use other tracking methods (like bounding box or mask) without implementing a keypoint algorithm – were excluded. (4) If a study solely used marker-based motion capture or sensor-based tracking (and not marker-less pose-estimation), it was excluded.

## 2.3 Study selection

The study selection process summarized in the flow diagram (Figure 1). First, titles and abstracts of all retrieved records were screened to exclude those unrelated to our focus.

To assist with screening and ensure, we employed the machine learning tool ASReview (de Bruin et al., n.d.), which prioritized records based on the likelihood of relevance. Using this tool, we iteratively screened approximately 1,000 articles by title, abstract, and full-text (if needed). This process allowed us to quickly identify candidate studies; any reference that the AI model flagged as relevant was given careful consideration, to maintain sensitivity (recall: 0.95). This process was primarily automated using advanced AI method with minimal human intervention. After initial screening, we obtained the full texts of studies that were potentially eligible. Authors then reviewed each full text in detail to determine final inclusion.

For the analysis, studies included in the qualitative synthesis; were categorized into one of three groups based on its primary focus: (1) Tool-focused papers: Studies primarily centered on the development or validation of a pose-estimation tool or algorithm. These typically introduced new software, frameworks, or technical improvements for tracking and analyzing rodent posture/behavior. (2) Method-focused papers: Studies that applied pose-estimation to propose a specific experimental method or paradigm. These papers often aimed to demonstrate how pose tracking can enhance a particular behavioral test or an experimental setup using pose-estimation. (3) Study-focused papers: Studies that used pose-estimation within the context of addressing a biological or disease-related research interest.

All of the included studies outlined a solution leveraging marker-less pose-estimation. Despite similarities in their methodologies and overlapping theme, we tried to categorized them based on the best-fitting approach.

## 2.4 Data extraction

Data of interests were manually pulled by authors from each included article using a preformatted form in an Excel workbook. For the tool-development paper, recorded fields included tool name, primary function or novelty, publication date, computational, code availability, and reported performance metrics across. For Method and Study papers, we extracted information on the experimental design and context: the rodent species/strain and any disease model or condition, the behavioral tests or tasks conducted, the pose-estimation software or method used, the number of key points tracked, the performance of tools, and the main outcomes or findings related to pose. The compiled data are organized into category-specific tables (Supplementary Tables 1–3).

## 2.5 Risk of bias assessment

We evaluated each study's risk of bias to maintain the trustworthiness of the findings. Given that the selected studies involved rodent experiments and computation analyses, in accordance with the ROBINS-I-V2 framework (for non-randomized studies), the risk of bias assessment was performed. Due to the nature of the current study, certain bias criteria were not applicable, but we still noted them according to the guidelines. The seven domains of bias include: Bias due to confounding; Bias in selection of participants into the study; Bias in classification of interventions; Bias due to deviations from intended interventions; Bias due to missing data; Bias in measurement of outcomes; and Bias in selection of the reported result. Each study was evaluated across multiple domains and assigned a risk level of “low risk,” “high risk,” or “unclear risk,” of bias. The risk of bias findings was compiled into summary tables and visualized via the Robvis tool (McGuinness and Higgins, 2021). These figures display the proportion of low/moderate/serious/critical risk in each domain and provide an individual risk profile for each paper.

## 2.6 Data synthesis

No meta-analysis was performed considering the study designs. Or rather, due to lack of the standardized benchmark across the marker-less pose-estimation field. Instead, we created a qualitative, descriptive review of the findings, structuring our results to align with the review objectives and presenting summary tables and figures to illustrate the main idea. Specifically, we synthesized the findings through (1) narrative summaries for each category (Tool, Method, Study), detailing their shared themes, technological advancements, and primary results. (2) We generated a year-by-year publication chart for tool-development studies and for method-application studies to track adoption patterns over time. (3) We classified pose-estimation tools according to their primary functions to illustrate the field's functional diversity. (4) We analyzed the technical aspects across tools (such as network backbones) and alongside experimental variables in method and study papers (identifying the most frequently used behavioral assays and disease models in pose-estimation research). (5) Where applicable, we compared performance metrics (such as accuracy and processing speed) across tools where data permitted. Observed trends and gaps found are highlighted across the study.

# 3 Results

## 3.1 Study selection

The comprehensive search across the selected sources yielded a large number of references. The PRISMA flow diagram in Figure 1a traces the study selection pathway from the initial database search through each subsequent screening stage. After removing duplicates, a total of 16,412 unique records were screened using ASReview (de Bruin et al., n.d.). Figure 1b represents the chronological record and Figure 1c the rate of relevant studies discovered using ASReview tool, which is the reason to screen the

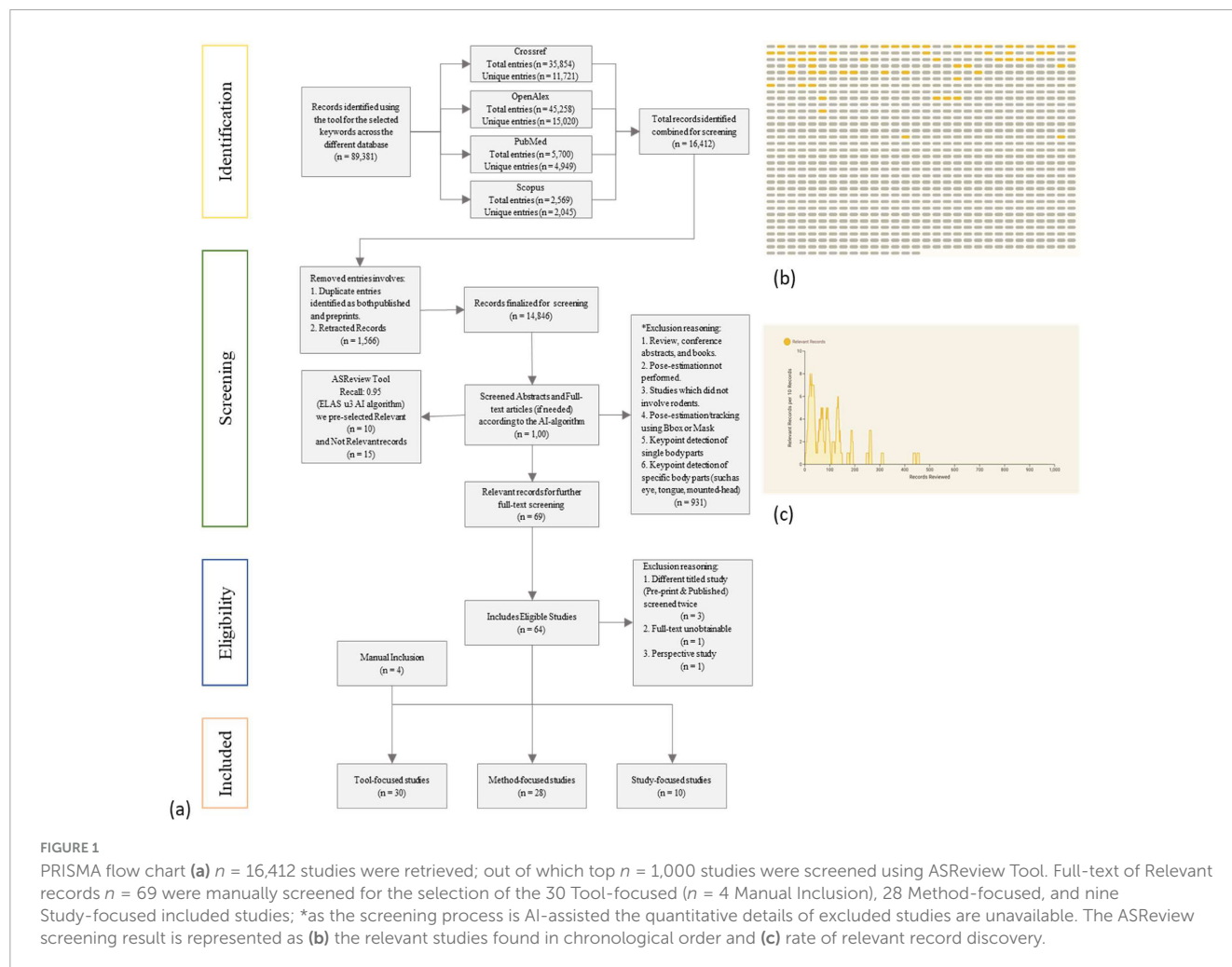


FIGURE 1

PRISMA flow chart (a)  $n = 16,412$  studies were retrieved; out of which top  $n = 1,000$  studies were screened using ASReview Tool. Full-text of Relevant records  $n = 69$  were manually screened for the selection of the 30 Tool-focused ( $n = 4$  Manual Inclusion), 28 Method-focused, and nine Study-focused included studies; \*as the screening process is AI-assisted the quantitative details of excluded studies are unavailable. The ASReview screening result is represented as (b) the relevant studies found in chronological order and (c) rate of relevant record discovery.

top 1,000 suggested records by title and abstract and full-text if needed. As shown, the recall of relevant studies falls well within this range, indicating a low likelihood of missing key studies. Of those, identifying 69 that were found to be relevant to the current study. The relevant studies were then advanced to full-text screening find 63 that met our inclusion criteria and were analyzed qualitatively. We classified these into three groups: 26 Tool-focused papers, 28 Method-focused papers, and nine Study-focused papers. In addition, we have included four Tool-focused papers manually, which were overlooked during screening.

### 3.2 Characteristics of the included studies

The key characteristics of all 67 included studies were summarized across three tables corresponding to Tool, Method, and Study. In the following sections, we detail each table's contents and discuss the key findings. Notably, each study confirms the added impact of pose-estimation in research on disease models.

Tool-focused studies ( $N = 30$ ): Table 1 provides an overview of the studies focused on keypoint or behavior detection employing marker-less pose-estimation techniques. These studies predominantly introduce new pose-estimation frameworks or

significant extensions to existing ones. The entries in list each tool's given name, the publication year, and running title is mentioned.

Method-focused studies ( $N = 28$ ): Table 2 summarizes the studies that we categorized as Method papers. These works employed existing pose-estimation techniques to advance or refine a particular experimental method in behavioral research. For each study, we list the name of method, behavioral test that was the focus (such as balance beam walking test, open field test, elevated plus maze, operant conditioning task, etc.), year of publication, and the running title.

Study-focused studies ( $N = 9$ ): Table 3 provides details on the included studies that we classified as Study papers, meaning their primary aim was to answer a biological or disease-related question, with pose-estimation being a means to that end. Each entry in table describes the disease or model under investigation, the behavioral assays used, year of publication, and the running title. All records included in the current studies performed the significant role in the advancement of the pose-estimation field.

### 3.3 Risk of bias assessment

We evaluated the risk of bias for the included studies, and a summary is presented in Supplementary Figures 1–3.

TABLE 1 List of selected studies in the category tool-focused ( $n = 30$ ).

Name of tool	Year	Title
DeepLabCut (Mathis et al., 2018)	2018	DeepLabCut: marker-less pose-estimation of user-defined body parts with deep learning
LEAP (Pereira et al., 2019)	2018	Fast animal pose-estimation using deep neural networks
DeepPoseKit (Graving et al., 2019)	2019	DeepPoseKit, a software toolkit for fast and robust animal pose estimation using deep learning
SimBa (Nilsson et al., 2020)	2020	Simple Behavioral Analysis (SimBA): an open source toolkit for computer classification of complex social behaviors in experimental animals
Anipose (Karashchuk et al., 2021)	2021	Anipose: A toolkit for robust markerless 3D pose estimation
CAPTURE (Marshall et al., 2021)	2021	Continuous Whole-Body 3D Kinematic Recordings across the Rodent Behavioral Repertoire
DANNCE (Dunn et al., 2021)	2021	Geometric deep learning enables 3D kinematic profiling across species and environments
DeepLabStream (Schweihoff et al., 2021)	2021	DeepLabStream enables closed-loop behavioral experiments using deep learning-based marker-less, real-time posture detection
MARS (Segalin et al., 2021)	2021	The Mouse Action Recognition System (MARS) software pipeline for automated analysis of social behaviors in mice
MouseVenue3D (Han et al., 2022)	2021	MouseVenue3D: A Marker-less Three-Dimension Behavioral Tracking System for Matching Two-Photon Brain Imaging in Free-Moving Mice
OptiFlex (Liu et al., 2021)	2021	OptiFlex: Multi-Frame Animal Pose-estimation Combining Deep Learning With Optical Flow
BehaviorDEPOT (Gabriel et al., 2022)	2022	BehaviorDEPOT is a simple, flexible tool for automated behavioral detection based on marker-less pose tracking
DeepLabCut (Lauer et al., 2022)	2022	Multi-animal pose-estimation, identification and tracking with DeepLabCut
SLEAP (Pereira et al., 2022)	2022	SLEAP: A deep learning system for multi-animal pose tracking
AlphaTracker (Chen et al., 2023)	2023	AlphaTracker: a multi-animal tracking and behavioral analysis tool
DeepOF (Bordes et al., 2023)	2023	Automatically annotated motion tracking identifies a distinct social behavioral profile following chronic social defeat stress
ContrastivePose (Zhou et al., 2023)	2023	ContrastivePose: A contrastive learning approach for self-supervised feature engineering for pose-estimation and behavioral classification of interacting animals
SaLSa (Sakata, 2023)	2023	SaLSa: A Combinatory Approach of Semi-Automatic Labeling and Long Short-Term Memory to Classify Behavioral Syllables
Temporal Semi-supervision method (Li et al., 2023)	2023	Improved 3D Marker-less Mouse Pose-estimation Using Temporal Semi-supervision
A-SOiD (Tillmann et al., 2024)	2024	A-SOiD, an active-learning platform for expert-guided, data-efficient discovery of behavior
ABNet (Chen et al., 2024)	2024	ABNet: AI-Empowered Abnormal Action Recognition Method for Laboratory Mouse Behavior
ARBEL (Barkai et al., 2024)	2024	ARBEL: A Machine Learning Tool with Light-Based Image Analysis for Automatic Classification of 3D Pain Behaviors
FABEL (Catto et al., 2024)	2024	FABEL: Forecasting Animal Behavioral Events with Deep Learning-Based Computer Vision
Keypoint-MoSeq (Weinreb et al., 2024)	2024	Keypoint-MoSeq: parsing behavior by linking point tracking to pose dynamics
Lightning Pose (Biderman et al., 2024)	2024	Lightning Pose: improved animal pose-estimation via semi-supervised learning, Bayesian ensembling, and cloud-native open-source tools

(Continued)



TABLE 1 (Continued)

Name of tool	Year	Title
REVEALS (Phadke et al., 2024)	2024	REVEALS: An Open Source Multi Camera GUI For Rodent Behavior Acquisition
STCS (Tang et al., 2024)	2024	Segmentation tracking and clustering system enables accurate multi-animal tracking of social behaviors
STPoseNet (Lv et al., 2024)	2024	STPoseNet: A real-time spatiotemporal network model for robust mouse pose-estimation
SuperAnimal Model (Ye et al., 2024)	2024	SuperAnimal pretrained pose-estimation models for behavioral analysis
Seizure Classification Pipeline (Yu et al., 2025)	2025	Integrating manual preprocessing with automated feature extraction for improved rodent seizure classification

In general, the methodological quality of studies varied, with many studies showing some risk of bias or reporting limitations in one or more domains. For the subset of studies, common issues included the lack of explicit randomization of animals into experimental groups and blinding. For example, several papers did not clearly state whether the experimenters were blinded to treatment or genotype during behavioral assessments, raising the risk of detection bias. On the other hand, most studies clearly defined their objectives and reported results thoroughly, so reporting bias was generally low. Many studies did not report factors such as animal selection, age, housing conditions, sleep-wake cycles; recording sessions plan likely deemed irrelevant for tool performance. However, overlooking these variables introduces confounding and selection bias. Additionally, some behavior classification studies omitted keypoint detection accuracy, raising concerns of reporting bias. Depending on the context, these biases may impact the reliability of datasets or the development of new tools or methods. In the graphical representation, we provide an aggregate view using color-blind safe format: indicate a low risk, a moderate, a high risk, and unclear risk in each domain. Overall, while no study was excluded due to quality, the assessment suggests that around half of the studies had at least one domain with a potential issue, meaning results should be interpreted with that context in mind. Conversely, about half of the studies (especially some tool papers and well-designed experiments) adequately addressed most bias concerns. It also highlights an area for improvement: future studies, particularly those implementing pose-estimation in biological experiments, should consider a robust design practice to strengthen confidence in their findings.

## 4 Quality evaluation of the included studies

Beyond the basic characteristics and bias assessment, we conducted further analyses to evaluate the trends and qualities of the included studies. In the following subsections, we present these findings, which encompass the temporal trends in publications, the evolution of tool features, the contexts in which pose-estimation is applied, and the usage patterns of different pose-estimation software. Each subsection entertains a specific aspect of idea and is represented by a figure or table.

### 4.1 Chronological list of tools

We first examined the timeline of publication related to the rodent pose-estimation tools (Figure 2a). The analysis revealed a clear upward trend over the past several years. From 2018 through 2025, there has been a steady increase in the number of new tools published per year. In the initial period, only a handful of tools were introduced, like DLC (Mathis et al., 2018) and LEAP (Pereira et al., 2019). The pace picked up modestly around 2020–2021 and then surged markedly in 2023 and 2024. The year 2024, in particular, saw the highest influx, with seven distinct new tools reported in that year alone, according to our dataset. This suggests that the field of pose-estimation for rodent analysis is in a phase of rapid innovation.

Several factors likely contributed to this growth: the success and wide adoption of initial tools probably spurred further developments with the increasing computational resources and open-source frameworks have lowered the barrier to creating new specialized tools (Voulodimos et al., 2018). By early 2025, the trend appears to continue, with at least a new tool already published in the first part of 2025, the seizure classification pipeline (Yu et al., 2025). The chronological trend underscores that the technology landscape is evolving quickly, and researchers are actively working on new solutions to extend capabilities. This also means that researchers will get access to an expanding array of tools to choose from with the advancement in the pose-estimation technology.

### 4.2 Primary purpose

We categorized each pose-estimation tool (from the Tool-focused studies) by its primary intended purpose (Figure 2b). This classification helps illustrate the diversity of approaches and end goals among the tools in this field. We identified several broad categories of tool functionality: (1) Core Pose-estimation: Tools whose primary aim is accurate marker-less tracking of animal key points. Examples: DLC (Mathis et al., 2018), LEAP (Pereira et al., 2019), SLEAP (Pereira et al., 2022) (functional successor of LEAP), Lightning Pose (Biderman et al., 2024), STPoseNet (Lv et al., 2024). These focus on improving the accuracy, robustness, or efficiency of pose detection. (2) Multi-Animal Tracking: Tools designed to track multiple animals simultaneously and possibly maintain individual identities. Examples: AlphaTracker (Chen et al., 2023), MARS (Segalin et al., 2021), SLEAP (Pereira et al., 2022) (which also falls

TABLE 2 List of selected studies in the category method-focused ( $n = 28$ ).

Name of method	Behavior test	Year	Title
DLCAnalyzer (Sturman et al., 2020)	Open-field; Elevated Plus Maze; Forced Swim Test	2020	Deep learning-based behavioral analysis reaches human accuracy and is capable of outperforming commercial solutions
*Not Specified* (Lang et al., 2020)	Unrestricted cage exploration Rodent Beam Walk	2020	Detecting and quantifying ataxia-related motor impairments in rodents using marker-less motion tracking with deep neural networks
Marker-less 2D Kinematic Analysis (Sato et al., 2022)	Treadmill walking	2021	Marker-less analysis of hindlimb kinematics in spinal cord-injured mice through deep learning
PS-VAE (Whiteway et al., 2021)	Head-fixed tasks Open Field Test	2021	Partitioning variability in animal behavioral videos using semi-supervised variational autoencoders
Automated Deep Phenotyping Pipeline (Klibaite et al., 2022)	Open Field Test	2022	Deep phenotyping reveals movement phenotypes in mouse neurodevelopmental models
VAME (Luxem et al., 2022)	Open Field Test	2022	Identifying behavioral structure from deep variational embeddings of animal motion
*Not Specified* (O'Neill et al., 2022)	Single Pellet Retrieval Task	2022	Marker-less tracking enables distinction between strategic compensation and functional recovery after spinal cord injury
*Not Specified* (Sheppard et al., 2022)	Open Field Test	2022	Stride-level analysis of mouse open-field behavior using deep-learning-based pose-estimation
AMBER (Lapp et al., 2023)	Maternal behavior observation	2023	Automated maternal behavior during early life in rodents (AMBER) pipeline
Posture Analysis Workflow (Wan et al., 2023)	Rodent Beam Walk	2023	An integrated workflow for 2D and 3D posture analysis during vestibular system testing in mice
SEB3R (Chelini et al., 2023)	Open Field Test	2023	Automated Segmentation of the Mouse Body Language to Study Stimulus-Evoked Emotional Behaviors
Two-bottle choice assay (Kim et al., 2023)	Two-bottle choice licking assay in fasted vs. sated mice	2023	Evaluation of mouse behavioral responses to nutritive versus non-nutritive sugar using a deep learning-based 3D real-time pose-estimation system
*Not Specified* (Bogachev et al., 2023)	Open Field Test	2023	Video-based marker-free tracking and multi-scale analysis of mouse locomotor activity and behavioral aspects in an open field arena: A perspective approach to the quantification of complex gait disturbances associated with Alzheimer's disease
*Not Specified* (Li et al., 2022)	Overground runway gait Treadmill locomotion	2023	A Multimode Marker-less Gait Motion Analysis System Based on Lightweight Pose-estimation Networks
Air-Stepping (Mistretta et al., 2024)	Air Stepping in neonatal mouse	2024	Air-stepping in the neonatal mouse: a powerful tool for analyzing early stages of rhythmic limb movement development
Arthritis Pain Assessment System (Li et al., 2024)	Automated Arthritis Pain Assessment	2024	Combining dual-view fusion pose-estimation and multi-type motion feature extraction to assess arthritis pain in mice
BAS (Piotrowski et al., 2024)	Rodent Beam Walk	2024	Phenotypic analysis of ataxia in spinocerebellar ataxia type 6 mice using DeepLabCut
KineWheelSystem (Albrecht et al., 2024)	Paw placement tracking on the KineWheel	2024	KineWheel-DeepLabCut Automated Paw Annotation Using Alternating Stroboscopic UV and White Light Illumination
MoSoMoTr (Sterley et al., 2024)	Social interaction assays	2024	Marker-less Mouse Tracking for Social Experiments
SBeA (Han et al., 2024)	Free-social behavior	2024	Multi-animal 3D social pose-estimation, identification and behavior embedding with a few-shot learning framework
String Pulling Analysis Pipeline (Sandhu et al., 2024)	Standing string-pulling for reward	2024	Information-theory analysis of mouse string-pulling agrees with Fitts's Law: Increasing task difficulty engages multiple sensorimotor modalities in a dual oscillator behavior
*Not Specified* (Zahran et al., 2024)	Three Chamber Sociability and Social Novelty Test	2024	Deep learning-based scoring method of the three-chamber social behavior test in a mouse model of alcohol intoxication. A comparative analysis of DeepLabCut, commercial automatic tracking and manual scoring
*Not Specified* (Bidgood et al., 2024)	Rodent Beam Walk	2024	Automated procedure to detect subtle motor alterations in the balance beam test in a mouse model of early Parkinson's disease
ArguelloALab (Sanabria et al., 2025)	Operant Self-Administration (Cocaine IVSA)	2025	Analysis of Operant Self-administration Behaviors with Supervised Machine Learning: Protocol for Video Acquisition and Pose-estimation Analysis Using DeepLabCut and Simple Behavioral Analysis

(Continued)

TABLE 2 (Continued)

Name of method	Behavior test	Year	Title
ForestWalk (Tozzi et al., 2025)	Rodent Beam Walk	2025	Forestwalk: A Machine Learning Workflow Brings New Insights Into Posture and Balance in Rodent Beam Walking
InteBOMB (Zhai et al., 2025)	Open-Field; Elevated Plus Maze; Forced Swim Test	2025	InteBOMB: Integrating generic object tracking and segmentation with pose-estimation for animal behavior analysis
LBWT-AT (Ruiz-Vitte et al., 2025)	Rodent Beam Walk	2025	Ledged Beam Walking Test Automatic Tracker: Artificial intelligence-based functional evaluation in a stroke model
Touchscreen-based rodent Continuous Performance Test (rCPT) (Li et al., 2025)	Touchscreen-based rCPT	2025	Time-on-task-related decrements in performance in the rodent continuous performance test are not caused by physical disengagement from the task

TABLE 3 List of selected studies in the category Studies-focused (*n* = 9).

Condition	Behavior test	Year	Title
Parkinsonism (6-OHDA lesion) (Andreoli et al., 2021)	Abnormal involuntary movements rating	2021	Distinct patterns of dyskinetic and dystonic features following D1 or D2 receptor stimulation in a mouse model of parkinsonism
Mechanical pain hypersensitivity (Zhang et al., 2022)	Home-cage exploration	2022	Automated preclinical detection of mechanical pain hypersensitivity and analgesia
Photothrombotic ischemic stroke (Weber et al., 2022)	Gait analysis	2022	Deep learning-based behavioral profiling of rodent stroke recovery
Thermal pain hypersensitivity (Reddy et al., 2023)	Thermal-plate assay	2023	A Deep-Learning Driven Investigation of the Circuit Basis for Reflexive Hypersensitivity to Thermal Pain
Neuropathic pain (Norris et al., 2023)	Tail suspension assay	2023	Spared nerve injury causes motor phenotypes unrelated to pain in mice
Alzheimer's disease models (Miller et al., 2024)	Open-field exploration	2024	Machine learning reveals prominent spontaneous behavioral changes and treatment efficacy in humanized and transgenic Alzheimer's disease models
Rett syndrome (Mykins et al., 2024)	Pup retrieval	2024	Multidimensional Analysis of a Social Behavior Identifies Regression and Phenotypic Heterogeneity in a Female Mouse Model for Rett Syndrome
Dopamine depletion model (Yang et al., 2024)	Open-field locomotion	2024	Dopamine lesions alter the striatal encoding of single-limb gait
Buprenorphine exposure (Sharma et al., 2025)	Elevated zero maze	2025	Machine learning and confirmatory factor analysis show that buprenorphine alters motor and anxiety-like behaviors in male, female, and obese C57BL/6J mice

under core pose), and STCS (Tang et al., 2024) (spatio-temporal clustering for socials). These are crucial for social interaction studies or high-throughput settings with group-housed animals. (3) 3D Pose-estimation: These go beyond 2D to reconstruct animal poses in three dimensions, often for more complex motor or biomechanical studies such as Anipose (Karashchuk et al., 2021), CAPTURE (Marshall et al., 2021), DANNCE (Dunn et al., 2021), and MouseVenus3D (Han et al., 2022). (4) Real-Time Tracking: Marker-less pose-estimation detection for closed-loop experiments, delivers posture-dependent stimuli by estimating animal pose online with millisecond latency, DeepLabStream (Schweihoff et al., 2021) is one of the tool for that purpose. (5) Behavior Classification: Tools that integrate a layer of identifying specific behaviors or actions from the pose data. Examples: SimBA (Nilsson et al., 2020) (which uses pose features to classify behaviors), BehaviorDEPOT (Gabriel et al., 2022), and ARBEL (Barkai et al., 2024). (6) Behavior Prediction: the highly specialized tools which forecast, flag abnormal behaviors [ABNet (Chen et al., 2024)] or detect pain [ARBEL (Barkai et al., 2024)] using trained models and pose dynamics. Even the prediction of future locomotion trajectories from past movements like FABEL (Catto

et al., 2024). (7) Infrastructure/Frameworks: Foundation models for pose-estimation across species such as SuperAnimal (Ye et al., 2024).

The resulting distribution shows that core pose-estimation, multi-animal tracking, 3D pose-estimation, real-time tracking, behavior classification, behavior prediction, and infrastructure/frameworks are very prominent needs that many tools address. Overall, this analysis highlights that the tools are being developed with different end goals in mind.

### 4.3 Architectural approaches in tools

We evaluated the algorithmic and architectural approaches employed by the various pose-estimation tools (Table 4). Virtually all modern rodent pose-estimation tools leverage deep learning, but there are variations in network architecture and training strategies: A majority use Convolutional Neural Network (CNN) backbones originally developed for image recognition or human pose-estimation (Grinciunaite et al., 2016). For example, ResNet-50 is a common backbone (used in DLC and others), often coupled



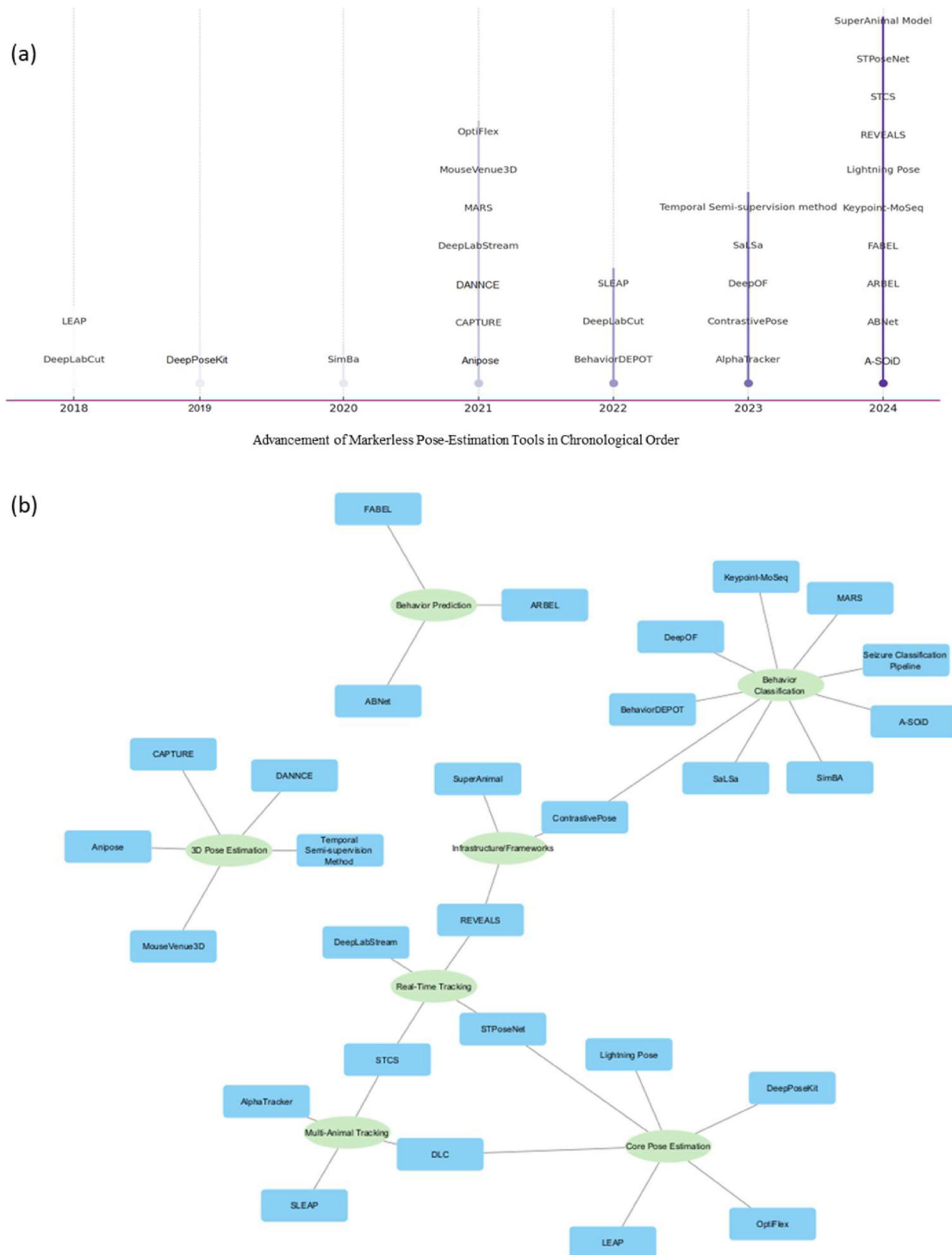


FIGURE 2

(a) Based on the year of publication, the timeline of marker-less rodent pose-estimation based keypoint detection and behavior detection tools. (b) The current network represents the primary intended purpose and key capabilities. The main categories (Green node) include: Core Pose-estimation, Multi-Animal Tracking, 3D Pose-estimation, Real-Time Tracking, Behavior Classification, Behavior Prediction, and Infrastructure/Frameworks. The published tools (Blue node) are linked with their intended purpose.

with deconvolution or upsampling layers to produce heatmaps for keypoint locations (Mathis et al., 2018). Some tools experimented with different backbones: AlphaTracker mentions DarkNet-53 and ResNet variants (Chen et al., 2023); LEAP used a variant of a stacked dense network (Pereira et al., 2019); similarly DeepPoseKit also uses

variant of a stacked dense network and stacked hourglass model (Graving et al., 2019); newer tools like STPoseNet may integrate spatial transformer networks (Lv et al., 2024). For multi-animal tracking, architectures often incorporate an identity association component. SLEAP (Pereira et al., 2022), for instance, can use

TABLE 4 The tools are categorized based on the architecture family.

Architecture family	Tools
DeepLabCut-based (ResNet-50)	ABNet (Chen et al., 2024) ARBEL (Barkai et al., 2024) BehaviorDEPOT (Gabriel et al., 2022) DeepLabCut (Mathis et al., 2018) DeepLabStream (Schweihoff et al., 2021) DeepOF (Bordes et al., 2023) Keypoint-MoSeq (Weinreb et al., 2024) REVEALS (Phadke et al., 2024) SaLSa (Sakata, 2023) SimBa (Nilsson et al., 2020)
YOLO-based	ContrastivePose (Zhou et al., 2023) Seizure Classification Pipeline (Yu et al., 2025) STCS (Tang et al., 2024)
YOLOv3-based	AlphaTracker (Chen et al., 2023)
YOLOv8-based	STPoseNet (Lv et al., 2024)
SLEAP-based (UNet/Hourglass)	FABEL (Catto et al., 2024) SLEAP (Pereira et al., 2022)
Semi-supervised Bayesian Ensemble	Lightning Pose (Biderman et al., 2024)
Pretrained ensemble (DeepLabCut Model)	SuperAnimal (Ye et al., 2024)
Other CNN	
Multi-view CNN	MouseVenue3D (Han et al., 2022)
Custom 15-layer CNN	LEAP (Pereira et al., 2019)
3D triangulation TCN	CAPTURE (Marshall et al., 2021)
CNN (FlexibleBaseline) Optical Flow	OptiFlex (Liu et al., 2021)
MSC-Multibox detector 8-stack Hourglass CNN	MARS (Segalin et al., 2021)
Custom 3D pose CNN	Temporal Semi-supervision Method (Li et al., 2023)

part affinity fields [similar to OpenPose (Cao et al., 2019)] or other graphical models to separate individuals. AlphaTracker’s pipeline combined a YOLO-based detection with pose-estimation, effectively splitting the task into detecting each animal and then finding key points (Chen et al., 2023).

4.4 Accuracy versus speed trade-offs

One practical consideration in pose-estimation tool performance is the trade-off between accuracy and speed. We aggregated the performance information reported in tool papers to qualitatively assess this trade-off. Different studies report performance in different ways, but two common measures for pose-estimation accuracy [often quantified by metrics like% of correct key points, mean pixel error, or mAP (mean average precision)] and runtime efficiency [measured in FPS (frames per second) processed, or whether the method can run in real-time]. We summarize schematically how tools tend to position themselves, available in Supplementary Tables 1–3. Generally, most tools cluster toward the high-accuracy end, given the emphasis on precise tracking in research. In the absence of

evaluation against a standardized benchmark dataset, the reported performance values in their studies are not directly comparable and should be interpreted as arbitrary. However, a subset extends toward the all-rounders as most of the tools are built upon DLC-based architecture.

4.5 Chronological list of methods

We also looked at the timeline of the Method-focused studies to see when researchers started incorporating pose-estimation into their experimental methods (Figure 3a). Unlike the tool development, which began in the 2018, the uptake of pose-estimation in general behavioral research shows a slight delay. The earliest Method category papers in our review were published around 2020. There was only one such study in 2020 that met our criteria and a couple in 2021. The count rises in 2022 and more sharply in 2023 and 2024, similar to the tool trend. Many published methods were focused on diverse behavioral assays such as gait analysis (Li et al., 2022; Sheppard et al., 2022), maternal care (Lapp et al., 2023), social interaction (Sterley et al., 2024; Zahran et al., 2024), and pain behavior (Li et al., 2024). Notably, 2024 emerged as a peak year, with a considerable number of new pipelines introduced. And the momentum appears to continue into 2025 (we already have five from early 2025 in our inclusion list). This timeline suggests that broad adoption by experimentalists lagged a year or two behind the introduction of major tools. It makes sense: tools like DLC (Mathis et al., 2018) became widely known around 2018–2019, and after some time for dissemination and training, more labs began applying them to their own experiments, leading to publications a year or more later. The accelerating trend in 2023–2024 indicates that pose-estimation is becoming more mainstream in the methods of behavioral labs. This analysis highlights the encouraging fact that the community is increasingly embracing these new methods, though it also points to a gap, it took a few years for many researchers to integrate these tools, suggesting a learning curve or initial resource barrier that needed to be overcome.

4.6 Pose-estimation tools used in method papers

We systematically reviewed the tools used for pose-estimation in both Method and Study category papers (37 papers total). As discussed earlier for method studies, DLC was the most commonly used tool (Figure 3b). DLC was used in 30 of 37 studies (~81%). This includes various versions (2D, 3D, multi-animal extensions) but collectively underscores its prevalence. Most of the method papers explicitly used DLC (Mathis et al., 2018) for their pose tracking, often in conjunction with downstream analysis frameworks such as SimBA (Nilsson et al., 2020). For example, tools like AMBER (Lapp et al., 2023), ArguelloALab (Sanabria et al., 2025), and BAS (Piotrowski et al., 2024) all incorporated DLC keypoints and used Random Forest classifiers for behavior annotation. Several tools also employed hybrid approaches by integrating pose-estimation with domain-specific algorithms. For instance, Air-Stepping employed circular statistics and EMG step-matching (Mistretta et al., 2024), while Posture Analysis Workflow

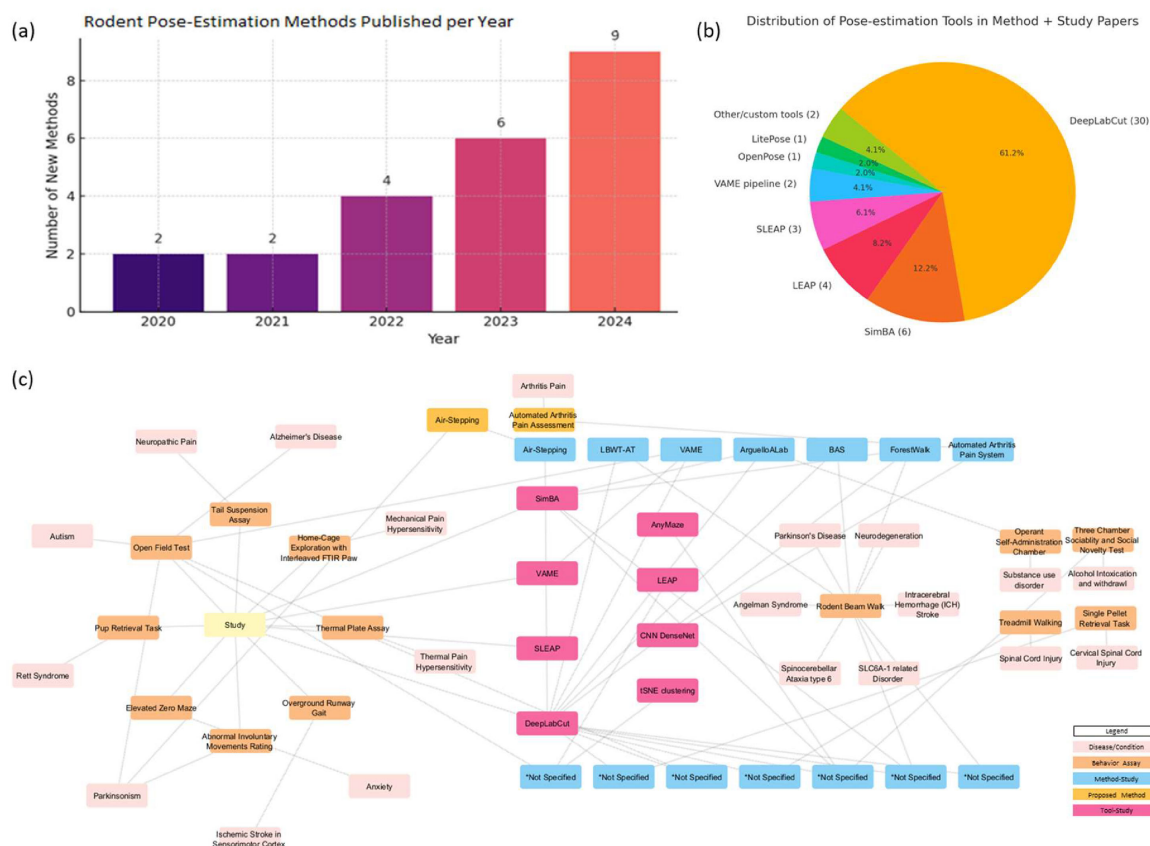


FIGURE 3

(a) Chronological milestone for method-focused studies. (b) Marker-less pose-estimation tool usage frequency across method development and disease studies. (c) The cross-analysis map of disease conditions, behavioral assays, study-focused or method-focused studies, and tools to present the current scenario of marker-less pose-estimation technology in the disease studies.

relied on FluoRender scripts for beam-walk kinematics (Wan et al., 2023). While performance reporting varied in detail, pixel error was the most common accuracy metric. Tools such as ForestWalk (Tozzi et al., 2025) and ArguelloALab (Sanabria et al., 2025) reported test errors ranging between 3–10 pixels, demonstrating practical precision for behavioral quantification in freely moving animals. DLC was the prominent framework found in the developments of the above-mentioned methods. Bias toward DLC may arise from its position as the earliest developed tool in the field development and established role as a benchmark tool in the field. And based on our findings, most of the behavior classification pipeline, and methodology are built upon DLC, using it as the foundational keypoint detection framework, followed by the application of specialized algorithms to address specific research problems. From a broader perspective, this indicates that researchers doing rodent experiments largely rely on a few well-established pose-estimation platforms rather than exploring the trend or writing their own from scratch.

LEAP (Pereira et al., 2019) was used in about four studies (mostly older ones before 2020) and SLEAP (Pereira et al., 2022) appearing in a few studies (approximately 3 out of 37). Those were typical cases needing multi-animal tracking or where authors were early adopters of this newer tool. No other single tool besides DLC and SLEAP showed up more than once or twice in the method and studies. This indicates that the community has largely coalesced

around one primary tool (DLC) for conducting pose-estimation in practice; the usage frequency we see reflects a lag: tools introduced a few years ago (like DLC, SLEAP) have usage now, whereas brand-new ones have little to no representation yet outside their own introduction papers. This metric also highlights a potential risk: with so much reliance on one tool, if that tool had any biases or limitations, many results could be affected similarly. Over time, we may see diversification as newer tools mature and demonstrate clear advantages. As of the data in our review, the pose-estimation landscape in practice is highly centered on DLC.

## 4.7 Behavioral assays addressed by method papers

We analyzed which behavioral tests or paradigms were most commonly addressed by the Method-focused studies (Table 5). This gives insight into where researchers find pose-estimation most useful in terms of types of behavior. The rodent balance beam walk (and similar gait/coordination tests) emerged as a frequently used paradigm in these papers. At least four independent method studies focused on beam walking tasks to evaluate motor coordination, often in the context of neurological disorders (Lang et al., 2020;

TABLE 5 Developed methods primarily focusing on disease conditions.

Behavioral domain	Behavioral assay	Method papers	Associated disease models
Locomotion and gait	Open Field Test	8	Alzheimer's disease (Bogachev et al., 2023; Miller et al., 2024)
	Rodent Beam Walk	5	Autism (Klibaite et al., 2022)
	Treadmill/Runway Walking	2	Angelman syndrome (Tozzi et al., 2025)
	Air-Stepping (Neonatal)	1	Spinocerebellar/Ataxia (Lang et al., 2020; Piotrowski et al., 2024)
	KineWheel Gait Task	1	Stroke (Weber et al., 2022; Ruiz-Vitte et al., 2025)
	Elevated Plus Maze	2	Parkinson's disease (Mistretta et al., 2024)
Anxiety-like behavior	Elevated Zero Maze	1	Spinal cord Injury (O'Neill et al., 2022; Sato et al., 2022)
			SLC6A-1 disorder (Tozzi et al., 2025)
			Anxiety-like Behavior (Sharma et al., 2025)
Rewards and operant conditioning	Operant/Reward Tasks	1	Cocaine self-administration (Sanabria et al., 2025)
	Three Chamber Sociability and Social Novelty Test	1	Alcohol consumption preference (Zahran et al., 2024)
Maternal care	Maternal Behavior	1	Maternal care deficits (Lapp et al., 2023)
Pain assessment	Arthritis Pain Gait Assay	1	Arthritis (Li et al., 2024)
Fine motor skills	Single-Pellet Reaching	1	Stroke Recovery (Weber et al., 2022)
	String-Pulling	1	Motor Deficit Models (Sandhu et al., 2024)
Spontaneous/longitudinal behavior	Homecage Observations	1	Stroke Recovery (Weber et al., 2022)

Tozzi et al., 2025) or injuries (Ruiz-Vitte et al., 2025). Pose-estimation is particularly well-suited here because it can count foot slips, measure crossing speed, and even detail how each paw moves – critical for detecting ataxia (Lang et al., 2020) or subtle motor deficits (Tozzi et al., 2025). Open field testing was another common assay, appearing either alone or in combination with other tests in several studies. In an open field, pose tracking gives not only total distance and speed (which could be done with simpler tracking) but also posture, limb movement patterns, and specific behaviors like rearing if 3D or multi-point is tracked (Klibaite et al., 2022). Some studies used open-field data to derive more complex metrics (such as unsupervised clustering of movement motifs) (Klibaite et al., 2022; Bogachev et al., 2023; Miller et al., 2024). Elevated Plus Maze and related anxiety tests (Sharma et al., 2025) (light-dark box, etc.) were present in a few papers. Pose-estimation here can automate measurements such as time spent in open vs. closed arms, as well as provide additional detail like head dips or stretch-attend postures if key points are tracked. Forced Swim Test (FST) was included in at least one study (Zhai et al., 2025). Social interaction tests and Operant behavior tests were featured in a couple of method papers. At least one method paper specifically dealt with an arthritis pain model (Li et al., 2024), using pose tracking to evaluate gait changes and pain-related behaviors like weight shifting. Another included a pain test in a broader context (Norris et al., 2023). A minority of studies looked at naturalistic behaviors (like maternal care or freely moving in homecage behaviors) for continuous pose tracking to capture subtle or long-term patterns (Lapp et al., 2023). This suggests that the community finds immediate value in applying pose tracking to tasks where movement is central, and deficits are quantitative. Other domains (social and cognitive tests) are less represented,

possibly because they are either harder to quantify or just emerging areas for such analysis. In time, as pose-estimation becomes more routine, we might see it applied even more broadly.

## 4.8 Disease models addressed by method papers

Among the Method-focused studies, a significant subset involved specific disease or injury models. We tallied the types of disease models featured (Table 5). Out of the 28 method papers, 14 (50%) incorporated an explicit disease or physiological challenge. The distribution of disease models in the method papers included: (1) Neurodegenerative and Neurological Disorders: Several studies focused on models of diseases such as Parkinson's disease (Yang et al., 2024), Alzheimer's disease (Bogachev et al., 2023; Miller et al., 2024), and spinocerebellar ataxia (Piotrowski et al., 2024). (2) Neurodevelopmental Disorders: Autism spectrum disorder models (Klibaite et al., 2022) and other developmental disorder models like Angelman syndrome (Tozzi et al., 2025) appeared in method studies. (3) Psychiatric/Addiction Models: Some papers involved substance use or withdrawal models [such as chronic alcohol exposure and withdrawal in mice (Zahran et al., 2024) and cocaine self-administration in rats (Sanabria et al., 2025)]. These studies used pose-estimation to observe changes in behavior, such as locomotor activity or specific actions during withdrawal periods. (4) Pain and Injury Models: Chronic pain models [like inflammatory arthritis in mice (Li et al., 2024)] and acute injury models [like spinal cord injury (Sato et al., 2022) or stroke (Weber et al., 2022)] were present. The stroke model study, for instance, introduced a new method to assess motor recovery by



tracking limb movement symmetry in a home cage monitoring (HCM) (Ruiz-Vitte et al., 2025), the rationale focuses on the idea of automated behavior analysis. From data presented it is evident that neurological disease models form the largest group among these method papers. This aligns with the intuition that motor deficits are a key feature of many neurological disorders, making those models a prime target for such methods.

## 4.9 Mapping behaviors to disease

The network is illustrated in Figure 3c represent the map linking behavioral assays to the disease models they were used to evaluate, across all relevant studies (both Method and Study categories). This mapping helps reveal if certain behaviors are particularly associated with certain types of disease research when using pose-estimation. We found that motor coordination tests like the beam walk are commonly used as mentioned above. These diseases naturally affect coordination, so researchers often employ beam walking or similar gait tests to quantify deficits (Bidgood et al., 2024; Ruiz-Vitte et al., 2025; Tozzi et al., 2025). Open field tests were used in a broad range of contexts, including Alzheimer's models, autism models, and as baseline in many others. Social interaction tests were specifically utilized in neurodevelopmental disorder studies and occasionally in neurodegeneration or psychiatric models. Tests commonly used in the domain of Anxiety research like elevated plus maze appeared in contexts like Alzheimer's disease (Miller et al., 2024) and in substance withdrawal studies (Sharma et al., 2025).

Pain-related gait assays (like automated scoring of limp or weight distribution) were obviously tied to pain models or nerve injury models (Norris et al., 2023). Complex behavior batteries were often used for models with uncertain phenotypes. For example, one comprehensive study on a Rett syndrome model mouse used a battery including open field, social test, and motor tests to capture a spectrum of behaviors via pose tracking (Mykins et al., 2024). The mapping highlights that each disease domain tends to utilize a relevant subset of behavioral tests, and pose-estimation is flexible enough to be applied to all these.

## 4.10 Limitation of screening strategy

The screening strategy leans extensively on automated processes from keyword searches to the retrieval of relevant publications. Publish or Perish software efficiently retrieves references, but it has limitations including a limit on results per query. Moreover, some retrieved entries lack abstracts, they might get prematurely excluded in subsequent screening step. ASReview relies on the structured abstracts and limits the inclusion of potentially relevant studies with missing or malformed metadata. Automated screening fell short when applied to extensive datasets. Notably, key studies such as Anipose (Karashchuk et al., 2021), A-SOiD (Tillmann et al., 2024), DANNCE (Dunn et al., 2021), and DeepPoseKit (Graving et al., 2019) were missed during the automated retrieval process.

## 5 Discussion

Our systematic review illustrates a period of rapid development and refinement of pose-estimation tools for rodent research. Over roughly 8 years (2017–2025), the field progressed from a handful of pioneering methods to a suite of sophisticated algorithms with diverse capabilities. Early tools established the feasibility of accurate marker-less tracking for instance, achieving sub-centimeter accuracy in detecting rodent limb positions effectively proving that computer vision could automate what was once manual scoring or marker-based motion capture (Mathis et al., 2018). Building on that foundation, successive tools have incrementally expanded the frontiers: introducing multi-animal tracking (Lauer et al., 2022) (to handle social groups or littermates in a cage), improving tracking speed (Phadke et al., 2024) (to approach real-time feedback, important for closed-loop experiments), and incorporating behavior classification and unsupervised analysis (Nilsson et al., 2020) (to interpret the raw pose data in terms of meaningful actions or patterns). The architectural evolution of these tools has been largely driven by advances in deep learning. Convolutional neural networks pre-trained on large datasets brought a step-change in keypoint detection accuracy around 2018, and since then, many tools have repurposed. The development trajectory of pose-estimation tools, from general frameworks like DLC (Mathis et al., 2018) and SLEAP (Pereira et al., 2022) to more specialized ones has been especially productive, developing a range of options that collectively cover many needs of the research community. The uptake of pose-estimation technology in preclinical behavioral experiments is clearly underway, as evidenced by the growing number of studies incorporating these methods. Our results show that starting around 2020, researchers began to apply pose tracking in classical rodent behavioral tests, and the trend has accelerated in the last few years. For example, in an open field test, instead of just recording the total distance travelled, researchers can quantify detailed trajectories, speed profiles, and even specific behavior postures automatically (Weber et al., 2022). In coordination tasks like the balance beam (Tozzi et al., 2025), subtle differences in how a rodent places its paws or maintains balance, which might be overlooked by human observation, are captured through automated video analysis (Ruiz-Vitte et al., 2025). This not only increases the sensitivity of the experiments (allowing detection of mild phenotypic differences or drug effects) but also their reliability [reducing observer bias and variability (Ruiz-Vitte et al., 2025)].

Our review highlights the use of pose-estimation to augment traditional behavioral assays rather than replace them outright. We see a methodological modernization in rodent behavioral science: the field is moving from stopwatches and manual counts toward automated, quantitative behavioral phenotyping. One of the most compelling findings of this review is how pose-estimation is empowering research on rodent models of disease. A transgenic mouse modeling early-stage Parkinson's disease might not exhibit obvious motor deficits under casual observation, but with pose-estimation, researchers can detect slight irregularities in gait or balance that herald disease onset (Mistretta et al., 2024). Such sensitive measures can serve as early biomarkers of disease progression or as endpoints to test intervention efficacy that would otherwise require a much larger sample sizes and sophisticated



experimentation to even notice. In pain and injury models, pose-estimation has allowed more objective and continuous monitoring of pain-related behaviors (Li et al., 2024). For instance, instead of relying solely on discrete scoring (like a pain scale based on observations at intervals), some studies continuously track how an animal shifts weight or adjusts posture, which can quantify pain levels with higher resolution over time (Norris et al., 2023). The ability to link specific behavioral metrics to disease conditions has much broader implications.

## 6 Limitation and future direction

While our review underscores many positive developments, it is important to acknowledge its limitations, as well as those in the field, and to suggest areas for future work. First, our review is limited by the scope of available literature, our search strategy, and the AI-assisted screening method. Being overly dependent on tools contains blind spots, which leads to the exclusion relevant studies as mentioned in the section above. And we focused on publications up to early 2016 and only those in English. It is possible that some relevant studies (especially very recent or non-indexed ones or those in other languages) were not captured. Second, the heterogeneity in study designs due to the lack of benchmarks across the included papers. This made it challenging to directly compare outcomes like accuracy or effect sizes. We relied on authors' reported metrics; some provided comprehensive evaluations, while others offered only qualitative assessments. A more standardized benchmarking across tools would greatly facilitate objective comparison. In our synthesis, we had to qualitatively assess trade-offs and trends rather than perform a quantitative meta-analysis due to this variability. Third, our categorization into tool-focused, method-focused, and study-focused studies was somewhat subjective and there is overlap between categories. Some tool papers also performed biological experiments to showcase their tool; some method papers introduced minor technical innovations. We chose categories to structure the review. This also reflects a limitation in the field: interdisciplinary studies sometimes defy neat categorization and valuable insights might be get over looked.

In terms of the research field's limitations revealed by this review, one notable aspect is the uneven adoption of pose-estimation. A large proportion of studies still come from either technology-oriented groups or early adopters instead of the preclinical research labs. Many traditional behavioral studies have yet to incorporate these methods. Barriers might include required expertise, computational resources, or simply inertia with established methods, as mentioned above. Another limitation is that while pose-estimation greatly improves measurement, it doesn't automatically interpret behavior; behavioral meaning must be inferred from pose data, and that still relies on expert knowledge or complementary experiments. Advanced analytics like machine learning classification or unsupervised clustering can help identify patterns, but there's a risk of over-reliance on algorithms without biological context. Future research should consider on linking pose-derived metrics more tightly to benchmark processes.

Looking ahead, future directions could include: (1) Expanding pose-estimation to more complex environments, most current

studies are in relatively controlled settings. Adaptive methods are required (Pereira et al., 2022). (2) Enhancing 3D pose-estimation for rodents; a few studies did this, but it's not widespread; improved 3D tracking could yield better readouts of complex behaviors (Lauer et al., 2022). (3) Integration with other data modalities such combining pose data with neural recordings, optogenetics triggers, or physiological readouts could provide a better understanding of behavior in context, and synchronized channel serve as a critical proof of concept, which are essential in establishing the benchmarks.

The advancement of pose-estimation has catalyzed progress in connected technologies, notably home cage monitoring (HCM) systems, which allow for 24/7, non-invasive collection of behavioral and physiological digital biomarkers (Baran et al., 2022). More recently, large language and vision-language models such as MouseGPT (Xu et al., 2025) and AmadeusGPT (Ye et al., 2023) have transformed behavior classification by directly interpreting raw video into open-vocabulary behavioral annotations also not relying on keypoint detection, steering the preclinical research toward new heights.

The current review's limitations are natural product of a new and very fast-moving field. We attempted to compile a comprehensive overview, and while some gaps remain, the trends identified are well defined. A continued push to upgrade these tools and broaden their adoption will help ensure that the insights from pose-estimation reach their full potential in advancing behavioral science.

## 7 Conclusion

Marker-less pose-estimation is the advent of advanced pose-estimation techniques, though this transformation is still in progress. We found that the development of new pose-estimation tools has been vigorous over the last several years, providing researchers with unprecedented capabilities to track and quantify behavior. The adoption of these tools in experimental studies is growing, particularly in areas where fine behavioral details matter, such as disease models and complex behavioral assays. Despite the availability of these tools, our review also highlights a persisting gap between technological advancements and their implement. Many rodent studies have yet to incorporate marker-less pose-estimation. And the majority of practitioners rely on a few key software tools. The evidence from the studies we synthesized indicates that embracing marker-less pose-estimation can significantly enhance the quality of data and conclusions in preclinical research. Whether it is validating a new therapy in a mouse model of disease or exploring fundamental questions of neuroscience, the ability to quantify behavior with high resolution leads to more robust and reproducible findings. In summary, the current trend in pose-estimation for rodent models is one of promising growth. Continued advancement is mostly likely to be the case. However, collaboration between tool developers and traditional researchers, for addressing practical barriers will be essential. By doing so, the field can ensure that the considerable advances in computational behavior analysis fully translate into deeper insights and breakthroughs in biomedical research.

## Data availability statement

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

## Author contributions

SB: Conceptualization, Data curation, Formal analysis, Investigation, Methodology, Visualization, Writing – original draft, Writing – review & editing. H-BK: Data curation, Formal analysis, Writing – review & editing. HSK: Validation, Writing – review & editing. BG: Data curation, Formal analysis, Writing – review & editing. J-IY: Conceptualization, Funding acquisition, Supervision, Validation, Writing – review & editing.

## Funding

The author(s) declare that financial support was received for the research and/or publication of this article. The current study was supported by a grant of National Research Foundation of Korea (NRF) funded by the Korea government Ministry of Science and ICT (MSIT) (No. RS-2021-NR060097). This research was supported by a grant of Korean ARPA-H Project through the Korea Health Industry Development Institute (KHIDI), funded by the Ministry of Health and Welfare, Republic of Korea (grant number: RS-2024-00507256).

## Acknowledgments

We would like to express our sincere gratitude to everyone who supported this work. We are also grateful to our colleagues and peers for their constructive feedback and encouragement. We have also employed AI-based tools to assist with text paraphrasing and grammar correction, which helped enhance the clarity and readability of the manuscript.

## References

- Albrecht, B., Schatz, A., Frei, K., and Winter, Y. (2024). KineWheel-DeepLabCut automated paw annotation using alternating stroboscopic UV and white light illumination. *ENEURO* 11:ENEURO.304–ENEURO.323. doi: 10.1523/ENEURO.0304-23.2024
- Andreoli, L., Abbaszadeh, M., Cao, X., and Cenci, M. A. (2021). Distinct patterns of dyskinetic and dystonic features following D1 or D2 receptor stimulation in a mouse model of parkinsonism. *Neurobiol. Dis.* 157:105429. doi: 10.1016/j.nbd.2021.105429
- Baran, S. W., Bratcher, N., Dennis, J., Gaburro, S., Karlsson, E. M., Maguire, S., et al. (2022). Emerging role of translational digital biomarkers within home cage monitoring technologies in preclinical drug discovery and development. *Front. Behav. Neurosci.* 15:758274. doi: 10.3389/fnbeh.2021.758274
- Barkai, O., Zhang, B., Turnes, B. L., Arab, M., Yarmolinsky, D. A., Zhang, Z., et al. (2024). ARBEL: A machine learning tool with light-based image analysis for automatic classification of 3D pain behaviors. *bioRxiv [Preprint]* doi: 10.1101/2024.12.01.625907
- Biderman, D., Whiteway, M. R., Hurwitz, C., Greenspan, N., Lee, R. S., Vishnubhotla, A., et al. (2024). Lightning pose: Improved animal pose estimation via semi-supervised learning, Bayesian ensembling and cloud-native open-source tools. *Nat. Methods* 21, 1316–1328. doi: 10.1038/s41592-024-02319-1
- Bidgood, R., Zubelzu, M., Ruiz-Ortega, J. A., and Morera-Herreras, T. (2024). Automated procedure to detect subtle motor alterations in the balance beam test in a mouse model of early Parkinson's disease. *Sci. Rep.* 14:862. doi: 10.1038/s41598-024-51225-1
- Bogachev, M., Sinitca, A., Grigorevichius, K., Pyko, N., Lyanova, A., Tsygankova, M., et al. (2023). Video-based marker-free tracking and multi-scale analysis of mouse locomotor activity and behavioral aspects in an open field arena: A perspective approach to the quantification of complex gait disturbances associated with Alzheimer's disease. *Front. Neuroinform.* 17:1101112. doi: 10.3389/fninf.2023.1101112
- Bordes, J., Miranda, L., Reinhardt, M., Narayan, S., Hartmann, J., Newman, E. L., et al. (2023). Automatically annotated motion tracking identifies a distinct social behavioral profile following chronic social defeat stress. *Nat. Commun.* 14:4319. doi: 10.1038/s41467-023-40040-3
- Cao, Z., Hidalgo, G., Simon, T., Wei, S.-E., and Sheikh, Y. (2019). OpenPose: Realtime multi-person 2D pose estimation using part affinity fields. *IEEE Trans. Pattern Anal. Mach. Intell.* 43, 172–186. doi: 10.1109/tpami.2019.2929257

## Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

## Generative AI statement

The authors declare that Generative AI was used in the creation of this manuscript. The author(s) verify and take full responsibility for the use of generative AI in the preparation of this manuscript. Generative AI was used we have employed AI-based tools to assist with text paraphrasing and grammar correction, which helped enhance the clarity and readability of the manuscript.

Any alternative text (alt text) provided alongside figures in this article has been generated by Frontiers with the support of artificial intelligence and reasonable efforts have been made to ensure accuracy, including review by the authors wherever possible. If you identify any issues, please contact us.

## Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

## Supplementary material

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fnbeh.2025.1663089/full#supplementary-material>

- Catto, A., O'Connor, R., Braunscheidel, K. M., Kenny, P. J., and Shen, L. (2024). FABEL: Forecasting animal behavioral events with deep learning-based computer vision. *bioRxiv [Preprint]* doi: 10.1101/2024.03.15.584610
- Chelini, G., Trombetta, E. M., Fortunato-Asquini, T., Ollari, O., Pecchia, T., and Bozzi, Y. (2023). Automated segmentation of the mouse body language to study stimulus-evoked emotional behaviors. *Eneuro* 10:ENEURO.514–ENEURO.522. doi: 10.1523/ENEURO.0514-22.2023
- Chen, Y., Guo, C., Han, Y., Hao, S., and Song, J. (2024). ABNet: AI-empowered abnormal action recognition method for laboratory mouse behavior. *Bioengineering* 11:930. doi: 10.3390/bioengineering11090930
- Chen, Z., Zhang, R., Fang, H.-S., Zhang, Y. E., Bal, A., Zhou, H., et al. (2023). AlphaTracker: A multi-animal tracking and behavioral analysis tool. *Front. Behav. Neurosci.* 17:111908. doi: 10.3389/fnbeh.2023.1111908
- de Bruin, J., Lombaers, P., Kaandorp, C., Teijema, J. J., van der Kuil, T., Yazan, B., et al. (n.d.). *ASReview LAB v2: Open-Source Text Screening with Multiple Agents and Oracles*. Available SSRN 5136987.
- Desland, F. A., Afzal, A., Warraich, Z., and Mocco, J. (2014). Manual versus automated rodent behavioral assessment: Comparing efficacy and ease of Bederson and Garcia neurological deficit scores to an open field video-tracking system. *J. Cent. Nerv. Syst. Dis.* 6, 7–14. doi: 10.4137/JCNSD.S13194
- Dubey, S., and Dixit, M. (2023). A comprehensive survey on human pose estimation approaches. *Multimed. Syst.* 29, 167–195. doi: 10.1007/s00530-022-00980-0
- Dunn, T., Marshall, J. D., Severson, K. S., Aldarondo, D., Hildebrand, D. G. C., Chettih, S. N., et al. (2021). Geometric deep learning enables 3D kinematic profiling across species and environments. *Nat. Methods* 18, 564–573. doi: 10.1038/s41592-021-01106-6
- Gabriel, C. J., Zeidler, Z., Jin, B., Guo, C., Goodpaster, C. M., Kashay, A. Q., et al. (2022). BehaviorDEPOT is a simple, flexible tool for automated behavioral detection based on markerless pose tracking. *Elife* 11:e74314. doi: 10.7554/eLife.74314
- Graving, J. M., Chae, D., Naik, H., Li, L., Koger, B., Costelloe, B. R., et al. (2019). DeepPoseKit, a software toolkit for fast and robust animal pose estimation using deep learning. *Elife* 8:e47994. doi: 10.7554/eLife.47994
- Grinciunaite, A., Gudi, A., Tasli, E., and Den Uyl, M. (2016). “Human pose estimation in space and time using 3d cnn,” in *Proceedings of the European Conference on Computer Vision*, (Berlin: Springer), 32–39.
- Hagelskjær, F., Buch, A. G., and Krüger, N. (2018). “Does vision work well enough for industry?” in *Proceedings of the 13th international joint conference on computer vision, imaging and computer graphics theory and applications - (volume 4)*, January 27–29, 2018, VISIGRAPP (4: VISAPP), Funchal, 198–205.
- Han, Y., Chen, K., Wang, Y., Liu, W., Wang, Z., Wang, X., et al. (2024). Multi-animal 3D social pose estimation, identification and behaviour embedding with a few-shot learning framework. *Nat. Mach. Intell.* 6, 48–61. doi: 10.1038/s42256-023-00776-5
- Han, Y., Huang, K., Chen, K., Pan, H., Ju, F., Long, Y., et al. (2022). MouseVenue3D: A markerless three-dimension behavioral tracking system for matching two-photon brain imaging in free-moving mice. *Neurosci. Bull.* 38, 303–317. doi: 10.1007/s12264-021-00778-6
- Harzing, A. W. (2007). *Publish or perish software*. Available online at: <https://harzing.com/resources/publish-or-perish>
- Karashchuk, P., Rupp, K. L., Dickinson, E. S., Walling-Bell, S., Sanders, E., Azim, E., et al. (2021). Anipose: A toolkit for robust markerless 3D pose estimation. *Cell Rep.* 36:109730. doi: 10.1016/j.celrep.2021.109730
- Kim, J., Kim, D., Jung, W., and Suh, G. S. B. (2023). Evaluation of mouse behavioral responses to nutritive versus nonnutritive sugar using a deep learning-based 3D real-time pose estimation system. *J. Neurogenet.* 37, 78–83. doi: 10.1080/01677063.2023.2174982
- Klibaite, U., Kislin, M., Verpeut, J. L., Bergeler, S., Sun, X., Shaevitz, J. W., et al. (2022). Deep phenotyping reveals movement phenotypes in mouse neurodevelopmental models. *Mol. Autism* 13:12. doi: 10.1186/s13229-022-00492-8
- Lang, J., Haas, E., Hübener-Schmid, J., Anderson, C. J., Pulst, S. M., Giese, M. A., et al. (2020). “Detecting and quantifying ataxia-related motor impairments in rodents using markerless motion tracking with deep neural networks,” in *Proceedings of the 2020 42nd Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC)*, (Piscataway, NJ: IEEE), 3642–3648. doi: 10.1109/EMBC44109.2020.9176701
- Lapp, H. E., Salazar, M. G., and Champagne, F. A. (2023). Automated maternal behavior during early life in rodents (AMBER) pipeline. *Sci. Rep.* 13:18277. doi: 10.1038/s41598-023-45495-4
- Lauer, J., Zhou, M., Ye, S., Menegas, W., Schneider, S., Nath, T., et al. (2022). Multi-animal pose estimation, identification and tracking with DeepLabCut. *Nat. Methods* 19, 496–504. doi: 10.1038/s41592-022-01443-0
- Li, H., Deng, Z., Yu, X., Lin, J., Xie, Y., Liao, W., et al. (2024). Combining dual-view fusion pose estimation and multi-type motion feature extraction to assess arthritis pain in mice. *Biomed. Signal Process. Control* 92:106080. doi: 10.1016/j.bspc.2024.106080
- Li, T., Severson, K. S., Wang, F., and Dunn, T. W. (2023). Improved 3D markerless mouse pose estimation using temporal semi-supervision. *Int. J. Comput. Vis.* 131, 1389–1405. doi: 10.1007/s11263-023-01756-3
- Li, W., Wang, S., Lei, J., Wang, X., Wang, L., Chen, K., et al. (2022). “A multimode markerless gait motion analysis system based on lightweight pose estimation networks,” in *Proceedings of the 2022 IEEE biomedical circuits and systems conference (BioCAS)*, (Piscataway, NJ: IEEE), 694–698.
- Li, Y., van Kralingen, T., Masi, M., Villanueva Sanchez, B., Mitchell, B., Johnson, J., et al. (2025). Time-on-task-related decrements in performance in the rodent continuous performance test are not caused by physical disengagement from the task. *NPP—Digital Psychiatry Neurosci.* 3:4. doi: 10.1038/s44277-025-00025-0
- Liu, X., Yu, S., Flierman, N. A., Loyola, S., Kamermans, M., Hoogland, T. M., et al. (2021). OptiFlex: Multi-frame animal pose estimation combining deep learning with optical flow. *Front. Cell. Neurosci.* 15:621252. doi: 10.3389/fncel.2021.621252
- Luxem, K., Mocellin, P., Fuhrmann, F., Kürsch, J., Miller, S. R., Palop, J. J., et al. (2022). Identifying behavioral structure from deep variational embeddings of animal motion. *Commun. Biol.* 5:1267. doi: 10.1038/s42003-022-04080-7
- Lv, S., Wang, J., Chen, X., and Liao, X. (2024). STPoseNet: A real-time spatiotemporal network model for robust mouse pose estimation. *Iscience* 27:109772. doi: 10.1016/j.isci.2024.109772
- Marshall, J. D., Aldarondo, D. E., Dunn, T. W., Wang, W. L., Berman, G. J., and Ölvéczky, B. P. (2021). Continuous whole-body 3D kinematic recordings across the rodent behavioral repertoire. *Neuron* 109, 420–437. doi: 10.1016/j.neuron.2020.11.016
- Mathis, A., Mamidanna, P., Cury, K. M., Abe, T., Murthy, V. N., Mathis, M. W., et al. (2018). DeepLabCut: Markerless pose estimation of user-defined body parts with deep learning. *Nat. Neurosci.* 21, 1281–1289. doi: 10.1038/s41593-018-0209-y
- McGuinness, L. A., and Higgins, J. P. T. (2021). Risk-of-bias VISualization (robvis): An R package and Shiny web app for visualizing risk-of-bias assessments. *Res. Synth. Methods* 12, 55–61. doi: 10.1002/jrsm.1411
- Miller, A. L., Flecknell, P. A., Leach, M. C., and Roughan, J. V. (2011). A comparison of a manual and an automated behavioural analysis method for assessing post-operative pain in mice. *Appl. Anim. Behav. Sci.* 131, 138–144. doi: 10.1016/j.applanim.2011.02.007
- Miller, S. R., Luxem, K., Lauderdale, K., Nambiar, P., Honma, P. S., Ly, K. K., et al. (2024). Machine learning reveals prominent spontaneous behavioral changes and treatment efficacy in humanized and transgenic Alzheimer's disease models. *Cell Rep.* 43:114870. doi: 10.1016/j.celrep.2024.114870
- Mistretta, O. C., Wood, R. L., English, A. W., and Alvarez, F. J. (2024). Air-stepping in the neonatal mouse: A powerful tool for analyzing early stages of rhythmic limb movement development. *J. Neurophysiol.* 131, 321–337. doi: 10.1152/jn.00227.2023
- Mykins, M., Bridges, B., Jo, A., and Krishnan, K. (2024). Multidimensional analysis of a social behavior identifies regression and phenotypic heterogeneity in a female mouse model for Rett syndrome. *J. Neurosci.* 44:e107823.2023. doi: 10.1523/JNEUROSCI.1078-23.2023
- Nassauer, A., and Legewie, N. M. (2019). Analyzing 21st century video data on situational dynamics—issues and challenges in video data analysis. *Soc. Sci.* 8:100. doi: 10.3390/socsci8030100
- Nilsson, S. R. O., Goodwin, N. L., Choong, J. J., Hwang, S., Wright, H. R., Norville, Z. C., et al. (2020). Simple behavioral analysis (SimBA)—an open source toolkit for computer classification of complex social behaviors in experimental animals. *bioRxiv [Preprint]* doi: 10.1101/2020.04.19.049452
- Norris, M. R., Dunn, S. S., Aravamuthan, B. R., and McCall, J. G. (2023). Spared nerve injury causes motor phenotypes unrelated to pain in mice. *bioRxiv [Preprint]* doi: 10.1101/2023.07.07.548155
- O'Neill, N., Mah, K. M., Badillo-Martinez, A., Jann, V., Bixby, J. L., and Lemmon, V. P. (2022). Markerless tracking enables distinction between strategic compensation and functional recovery after spinal cord injury. *Exp. Neurol.* 354:114085. doi: 10.1016/j.expneurol.2022.114085
- Pereira, T. D., Aldarondo, D. E., Willmore, L., Kislin, M., Wang, S. S.-H., Murthy, M., et al. (2019). Fast animal pose estimation using deep neural networks. *Nat. Methods* 16, 117–125. doi: 10.1038/s41592-018-0234-5
- Pereira, T. D., Tabris, N., Matsliah, A., Turner, D. M., Li, J., Ravindranath, S., et al. (2022). SLEAP: A deep learning system for multi-animal pose tracking. *Nat. Methods* 19, 486–495. doi: 10.1038/s41592-022-01426-1
- Phadke, R. A., Wetzel, A. M., Fournier, L. A., Brack, A., Sha, M., Padró-Luna, N. M., et al. (2024). REVEALS: An open-source multi-camera GUI for rodent behavior acquisition. *Cereb. Cortex* 34:bhae421. doi: 10.1093/cercor/bhae421
- Piotrowski, D., Clemensson, E. K. H., Nguyen, H. P., and Mark, M. D. (2024). Phenotypic analysis of ataxia in spinocerebellar ataxia type 6 mice using DeepLabCut. *Sci. Rep.* 14:8571. doi: 10.1038/s41598-024-59187-0
- Reddy, P., Vasudeva, J., Shah, D., Prajapati, J. N., Harikumar, N., and Barik, A. (2023). A deep-learning driven investigation of the circuit basis for reflexive hypersensitivity to thermal pain. *Neuroscience* 530, 158–172. doi: 10.1016/j.neuroscience.2023.08.023

- Reed, P., Holdaway, K., Isensee, S., Buie, E., Fox, J., Williams, J., et al. (1999). User interface guidelines and standards: progress, issues, and prospects. *Interact. Comput.* 12, 119–142. doi: 10.1016/S0953-5438(99)00008-9
- Ruiz-Vitte, A., Gutiérrez-Fernández, M., Laso-García, F., Piniella, D., Gómez-de Frutos, M. C., Díez-Tejedor, E., et al. (2025). Ledge Beam walking test automatic tracker: Artificial intelligence-based functional evaluation in a stroke model. *Comput. Biol. Med.* 186:109689. doi: 10.1016/j.compbiomed.2025.109689
- Sakata, S. (2023). SaLSa: A combinatory approach of semi-automatic labeling and long short-term memory to classify behavioral syllables. *Eneuro* 10:ENEURO.201–ENEURO.223. doi: 10.1523/ENEURO.0201-23.2023
- Sanabria, L. F. P., Voutour, L. S., Kaufman, V. J., Reeves, C. A., Bal, A. S., Maureira, F., et al. (2025). Analysis of operant self-administration behaviors with supervised machine learning: protocol for video acquisition and pose estimation analysis using deeplabcut and simple behavioral analysis. *Eneuro* 12:ENEURO.31–ENEURO.24. doi: 10.1523/ENEURO.0031-24.2024
- Sandhu, P. S., Agha, B. M., Inayat, S., Singh, S., Ryait, H. S., Mohajerani, M. H., et al. (2024). Information-theory analysis of mouse string-pulling agrees with Fitts's Law: Increasing task difficulty engages multiple sensorimotor modalities in a dual oscillator behavior. *Behav. Brain Res.* 456, 114705. doi: 10.1016/j.bbr.2023.114705
- Sato, Y., Kondo, T., Shinozaki, M., Shibata, R., Nagoshi, N., Ushiba, J., et al. (2022). Markerless analysis of hindlimb kinematics in spinal cord-injured mice through deep learning. *Neurosci. Res.* 176, 49–56. doi: 10.1016/j.neures.2021.09.001
- Schweihoff, J. F., Loshakov, M., Pavlova, I., Kück, L., Ewell, L. A., and Schwarz, M. K. (2021). DeepLabStream enables closed-loop behavioral experiments using deep learning-based markerless, real-time posture detection. *Commun. Biol.* 4:130. doi: 10.1038/s42003-021-01654-9
- Segalin, C., Williams, J., Karigo, T., Hui, M., Zelikowsky, M., Sun, J. J., et al. (2021). The mouse action recognition system (MARS) software pipeline for automated analysis of social behaviors in mice. *Elife* 10:e63720. doi: 10.7554/eLife.63720
- Sharma, O., Mykins, M., Bergee, R. E., Price, J. M., O'Neil, M. A., Mickels, N., et al. (2025). Machine learning and confirmatory factor analysis show that buprenorphine alters motor and anxiety-like behaviors in male, female, and obese C57BL/6J mice. *J. Neurophysiol.* 133, 502–512. doi: 10.1152/jn.00507.2024
- Sheppard, K., Gardin, J., Sabnis, G. S., Peer, A., Darrell, M., Deats, S., et al. (2022). Stride-level analysis of mouse open field behavior using deep-learning-based pose estimation. *Cell Rep.* 38:110231. doi: 10.1016/j.celrep.2021.110231
- Sterley, T.-L., Cheng, N., Bains, J. S., and Murari, K. (2024). Markerless mouse tracking for social experiments. *Eneuro* 11:ENEURO.154–ENEURO.122. doi: 10.1523/ENEURO.0154-22.2023
- Sturman, O., von Ziegler, L., Schläppi, C., Akyol, F., Privitera, M., Slominski, D., et al. (2020). Deep learning-based behavioral analysis reaches human accuracy and is capable of outperforming commercial solutions. *Neuropsychopharmacology* 45, 1942–1952. doi: 10.1038/s41386-020-0776-y
- Tang, C., Zhou, Y., Zhao, S., Xie, M., Zhang, R., Long, X., et al. (2024). Segmentation tracking and clustering system enables accurate multi-animal tracking of social behaviors. *Patterns* 5:101057. doi: 10.1016/j.patter.2024.101057
- Tillmann, J. F., Hsu, A. I., Schwarz, M. K., and Yttri, E. A. (2024). A-SOiD, an active-learning platform for expert-guided, data-efficient discovery of behavior. *Nat. Methods* 21, 703–711. doi: 10.1038/s41592-024-02200-1
- Tozzi, F., Zhang, Y., Narayanan, R., Roqueiro, D., and O'Connor, E. C. (2025). Forestwalk: A machine learning workflow brings new insights into posture and balance in rodent beam walking. *Eur. J. Neurosci.* 61:e70033. doi: 10.1111/ejn.70033
- Voulodimos, A., Doulamis, N., Doulamis, A., and Protopapadakis, E. (2018). Deep learning for computer vision: A brief review. *Comput. Intell. Neurosci.* 2018:7068349. doi: 10.1155/2018/7068349
- Wan, Y., Edmond, M. A., Kitz, C., Southern, J., and Holman, H. A. (2023). An integrated workflow for 2D and 3D posture analysis during vestibular system testing in mice. *Front. Neurol.* 14:1281790. doi: 10.3389/fneur.2023.1281790
- Weber, R. Z., Mulders, G., Kaiser, J., Tackenberg, C., and Rust, R. (2022). Deep learning-based behavioral profiling of rodent stroke recovery. *BMC Biol.* 20:232. doi: 10.1186/s12915-022-01434-9
- Weinreb, C., Pearl, J. E., Lin, S., Osman, M. A. M., Zhang, L., Annappagada, S., et al. (2024). Keypoint-MoSeq: Parsing behavior by linking point tracking to pose dynamics. *Nat. Methods* 21, 1329–1339. doi: 10.1038/s41592-024-02318-2
- Whiteway, M. R., Biderman, D., Friedman, Y., Dipoppa, M., Buchanan, E. K., Wu, A., et al. (2021). Partitioning variability in animal behavioral videos using semi-supervised variational autoencoders. *PLoS Comput. Biol.* 17:e1009439. doi: 10.1371/journal.pcbi.1009439
- Xu, T., Zhou, T., Wang, Y., Yang, P., Tang, S., Shao, K., et al. (2025). MouseGPT: A large-scale vision-language model for mouse behavior analysis. *arXiv [Preprint]* doi: 10.48550/arXiv.2503.10212
- Yang, L., Singla, D., Wu, A. K., Cross, K. A., and Masmanidis, S. C. (2024). Dopamine lesions alter the striatal encoding of single-limb gait. *Elife* 12:R92821. doi: 10.7554/eLife.92821
- Ye, S., Filippova, A., Lauer, J., Schneider, S., Vidal, M., Qiu, T., et al. (2024). SuperAnimal pretrained pose estimation models for behavioral analysis. *Nat. Commun.* 15:5165. doi: 10.1038/s41467-024-48792-2
- Ye, S., Lauer, J., Zhou, M., Mathis, A., and Mathis, M. (2023). “AmadeusGPT: A natural language interface for interactive animal behavioral analysis,” in *Thirty-seventh Conference on Neural Information Processing Systems (NeurIPS) 2023*. Available online at: <https://arxiv.org/abs/2307.04858>
- Yu, A., Singh, M., Pandey, A., Dybas, E., Agarwal, A., Kao, Y., et al. (2025). Integrating manual preprocessing with automated feature extraction for improved rodent seizure classification. *Epilepsy Behav.* 165:110306. doi: 10.1016/j.yebeh.2025.110306
- Zahrán, M. A., Manas-Ojeda, A., Navarro-Sánchez, M., Castillo-Gómez, E., and Olucha-Bordonau, F. E. (2024). Deep learning-based scoring method of the three-chamber social behaviour test in a mouse model of alcohol intoxication. A comparative analysis of DeepLabCut, commercial automatic tracking and manual scoring. *Heliyon* 10:e36352. doi: 10.1016/j.heliyon.2024.e36352
- Zhai, H., Yan, H.-Y., Zhou, J.-Y., Liu, J., Xie, Q.-W., Shen, L.-J., et al. (2025). InteBOMB: Integrating generic object tracking and segmentation with pose estimation for animal behavior analysis. *Zool. Res.* 46, 355–369. doi: 10.24272/j.issn.2095-8137.2024.268
- Zhang, Z., Roberson, D. P., Kotoda, M., Boivin, B., Bohnslav, J. P., González-Cano, R., et al. (2022). Automated preclinical detection of mechanical pain hypersensitivity and analgesia. *Pain* 163, 2326–2336. doi: 10.1097/j.pain.0000000000002680
- Zhou, T., Cheah, C. C. H., Chin, E. W. M., Chen, J., Farm, H. J., Goh, E. L. K., et al. (2023). ContrastivePose: A contrastive learning approach for self-supervised feature engineering for pose estimation and behavioral classification of interacting animals. *Comput. Biol. Med.* 165:107416. doi: 10.1016/j.compbiomed.2023.107416