Check for updates

Machine learning and deep learning frameworks for the automated analysis of pain and opioid withdrawal behaviors

Jacob R. Bumgarner*†, Darius D. Becker-Krail†,
Rhett C. White and Randy J. Nelson

Department of Neuroscience, Rockefeller Neuroscience Institute, West Virginia University,
Morgantown, WV, United States

The automation of behavioral tracking and analysis in preclinical research can serve to advance the rate of research outcomes, increase experimental scalability, and challenge the scientific reproducibility crisis. Recent advances in the efficiency, accuracy, and accessibility of deep learning (DL) and machine learning (ML) frameworks are enabling this automation. As the ongoing opioid epidemic continues to worsen alongside increasing rates of chronic pain, there are ever-growing needs to understand opioid use disorders (OUDs) and identify non-opioid therapeutic options for pain. In this review, we examine how these related needs can be advanced by the development and validation of DL and ML resources for automated pain and withdrawal behavioral tracking. We aim to emphasize the utility of these tools for automated behavioral analysis, and we argue that currently developed models should be deployed to address novel questions in the fields of pain and OUD research.

KEYWORDS

pain, opioid withdrawal, opioid use disorder (OUD), deep learning, machine learning, markerless tracking, automated behavioral analysis, pose estimation

Introduction

The ongoing opioid epidemic is currently the worst it has been in the history of the United States. Exacerbated by the global coronavirus disease 2019 (COVID-19) pandemic, in 2021 the U.S. saw the highest annual drug overdose death toll ever recorded, with an estimated 107,622 overdose deaths (Ahmad et al., 2021). Of these deaths, 75% involved opioid use (Ahmad et al., 2021). Among other factors, one major contribution to the opioid epidemic is the prevalence of chronic pain (Volkow and McLellan, 2016). Because of the lack of widespread non-opioid analgesics and alternative therapeutic options, opioids are continually prescribed for chronic pain management, despite the number of associated risk factors (Nadeau et al., 2021). As a result of this factor and increasing accessibility to potent synthetic opioids, including fentanyl, there

is an ever-growing need to identify treatment strategies that can mitigate this ongoing crisis (Skelly et al., 2020).

Translational models of pain and opioid withdrawal behaviors are crucial for the development of safer non-opioid analgesics and other treatment options for opioid use disorders (OUDs). A natural aspect of these preclinical experiments is the characterization of behavioral responses in translational models, particularly rodent models, which to date has primarily relied on human visual perception. As such, current behavioral paradigms in pain or opioid withdrawal research in rodents still rely on manual scoring (Deuis et al., 2017; Bravo et al., 2020, 2021). This process is often time-consuming, labor-intensive, and error-prone, even when conducted by trained researchers. The current approach of manual scoring also has limited scalability and introduces potential contributions to the reproducibility crisis.

Recent advances in the efficiency and accessibility of deep learning (DL) and machine learning (ML) frameworks have provided a clear route to challenge the issues associated with manual scoring of pain behavior (Mathis and Mathis, 2020). Widely accessible open-source tools being developed and used in preclinical research are now capable of automatically labeling and quantifying complex behaviors with accuracy that can match human performance (Table 1; Sturman et al., 2020b). By addressing the challenges associated with manual scoring, accurate automated behavioral analysis tools that leverage the utility of ML and DL frameworks can improve the rigor and throughput of preclinical and ecological research (Anderson and Perona, 2014). Moreover, these tools can provide improved insight into naturalistic behaviors that may otherwise go unnoticed by human observers (Datta et al., 2019). In this perspective, we outline the current ML and DL frameworks being used to label behavior in translational models and look ahead at how these tools could be applied to accelerate the preclinical study of pain and OUD.

Artificial intelligence as a tool for behavioral analysis

Automated behavior tracking in most experimental protocols generally follows a two-stage pipeline (Figure 1). The first stage of most of these pipelines involves video recording and subsequent point labeling/tracking of specific body parts of the studied animals. Numerous highly effective tools exist for point-tracking without added markers on the animals (i.e., markerless tracking). Examples of these tools include DeepLabCut (Mathis et al., 2018; Lauer et al., 2022), SLEAP (Pereira et al., 2022), DANNCE (Dunn et al., 2021), and Anipose (which relies on DeepLabCut) (Karashchuk et al., 2021). Broadly, these markerless point estimation tools use convolutional neural network (CNN) models with encoder-decoder architectures to create probability density plots for each trained feature. Each pixel value in the output plots represents the probability of the presence of the feature of interest. The probability densities are then used to localize points within the images. Following the identification of the desired points, whole-body or whole-limb point skeletons can be constructed for subsequent pose estimation and behavioral classifications. Moreover, several point estimation models can produce point labels and point skeletons for multiple animals within the same video recording (Lauer et al., 2022; Pereira et al., 2022).

The second stage of automated behavioral analysis involves the process of extracting meaningful information from the labeled point data (Sturman et al., 2020a). The approaches in this stage are much more varied and specific to the desired type of behavioral analysis being conducted. Many of the tools used in this stage of analysis rely on traditional tracking algorithms as well as supervised and unsupervised ML and DL. Basic behaviors, such as total movement, zone entries, and even paw flicks, can be tracked and quantified with traditional algorithms. In contrast, the identification of more abstract or nuanced behaviors, such as grooming, shaking, licking, and biting, require the use of more complicated algorithms [e.g., UMAP (McInnes et al., 2018)], ML algorithms (e.g., Random forest classifiers), and DL models (e.g., convolutional recurrent neural networks). Tools used in this domain to extract and quantify pose and behavioral information include SiMBA (Nilsson et al., 2020), B-SOiD (Hsu and Yttri, 2021), MoSeq (Wiltschko et al., 2015, 2020), and uBAM (Brattoli et al., 2021), among others (Sturman et al., 2020a; Hu et al., 2022).

Importantly, numerous tools and pipelines that can be used for both stages of automated behavioral analysis have been developed and validated for automated scoring of pain and opioid withdrawal behaviors. Moreover, several of these tools are open-source and easily accessible. In the next two sections, we will discuss the development, validity, and verification of several automated behavioral analysis tools in these two fields.

Automated analysis of pain behavior

Pain is a highly subjective experience that involves both affective and sensory-discriminative neurocircuitry engagement (Baliki and Apkarian, 2015). Preclinical research relies on the visual and sometimes acoustic observation of behavioral alterations in non-human animals to determine the presence of pain. Without using automated analysis techniques, pain scoring typically involves real-time visual inspections and timings of discrete behavioral events, such as paw flicking, extended paw withdrawals, jumping, grooming, licking/biting, and even squeaking (Deuis et al., 2017). Behavioral testing

TABLE 1   Summary of automated pain and opioid use disorder (OUD) behavioral analysis articles.

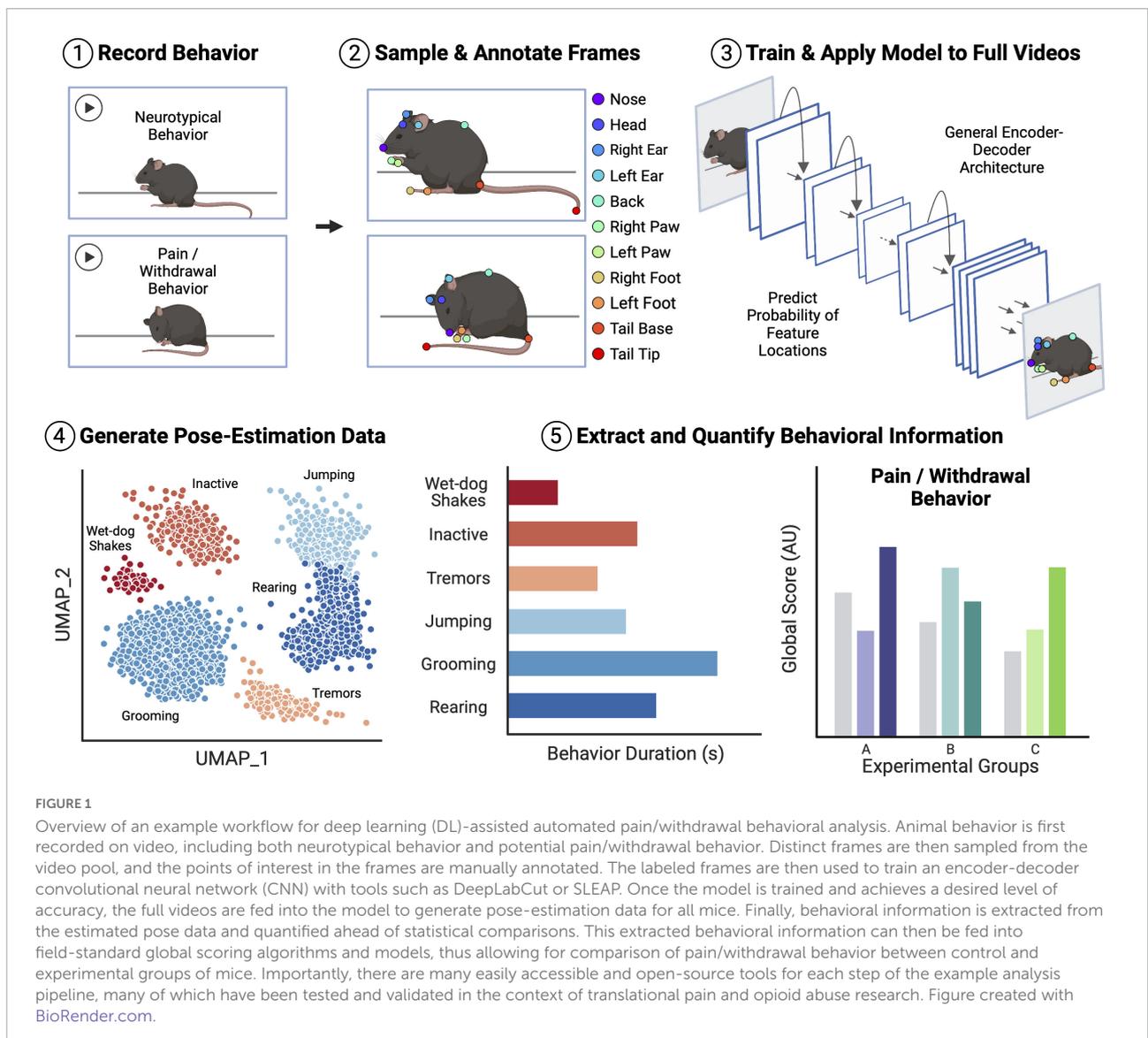| Citation	Relevant field	Behaviors tracked	Frameworks and software used	Repository and trained model links
Abdus-Saboor et al., 2019	Pain	Paw and face labeling	PainAssaySVM	https://github.com/longdecision/PainAssaySVM
Jones et al., 2020	Pain	Peak paw withdrawl height and guarding duration	SLEAP, PAWS, and ProAnalyst	https://github.com/crtwomey/paws
Bohic et al., 2021	Pain	Allodynia and others	DeepLabCut, B-SOiD, MoSeq, PAWS, pybasicbayes, scikit-learn, and Gensim	NA
Kopaczka et al., 2018	Pain	Mouse grimace scale	PyTorch	NA
Andresen et al., 2020	Pain	Postoperative pain	TensorFlow, OpenCV	NA
Tuttle et al., 2018	Pain	Pain or no pain	TensorFlow	NA
Kobayashi et al., 2021	Pain and OUD	Scratching	Keras, OpenCV	NA
Wotton et al., 2020	Pain and OUD	Licking/non-licking events such as paw flick	DeepLabCut, MatLab for GentleBoost model, and k-nearest neighbor classifier.	https://www.kumarlab.org/2020/10/05/machine-learning-based-automated-phenotyping-of-inflammatory-nocifensive-behavior-in-mice/
Kopaczka et al., 2019	Pain and OUD	General movement, grooming, and resting	NA	NA
Dao et al., 2021	Pain and OUD	Rat ultrasonic vocalizations	DeepSqueak	NA
Murphy et al., 2021	OUD	Jumping, rearing, grooming, tremors, etc.	DeepLabCut, SimBA,	NA

sessions are also frequently recorded for subsequent timing and count verifications or for delayed behavioral scoring.

Both real-time and recording-based behavioral scoring are highly subjective skills that require expert training and careful, rigorous attention. Real-time behavioral observations can suffer from experimenter fatigue, impaired viewing angles, and differing opinions over discrete events. Recording-based observations offer improvements over real-time scoring, as behavioral observations can be more carefully confirmed with repeated inspections, provided that the multiple viewing angles and high frame rates are used to prevent missed or obscured behavior by animal movement or crouching. However, this latter approach can involve long hours of staring at screens, which can lead to experimenter fatigue and potential inaccuracies. With careful recording setups and pipeline validation, automated analysis techniques offer the ability to overcome several of these issues (Fried et al., 2020). Indeed, numerous models have been developed to automate the analysis of scratching, paw withdrawal, general pain states, and mouse-grimace scale scoring.

Several automated models have been independently developed to identify scratching and hind paw flicking/withdrawal behaviors. One model used a convolutional recurrent neural network that relies on an extended temporal window of behavioral information to determine whether scratching was occurring in any given frame (Kobayashi et al., 2021). The authors reported an overall 94.8% accuracy of scratching detection and validated the automated detection

in a dinitrofluorobenzene-induced dermatitis experiment (Kobayashi et al., 2021). Another group used DeepLabCut to track paw movement from a bottom-up view and the GentleBoost classifier to label licking/non-licking behavioral events (Wotton et al., 2020). This model had a 98% accuracy in second-by-second behavioral classification, and the practical performance of this model was validated with analgesic and formalin injections (Wotton et al., 2020). Lastly, a third group developed an ML-based analysis pipeline that uses manual paw and face labeling to generate a rigorous pain score (Abdus-Saboor et al., 2019). The pipeline reduces paw and facial movement feature dimensionality with principal component analyses into principal component scores, which are fed into a trained support vector machine (SVM) that provides a single pain metric. Importantly, this approach could be further automated in the future by using point-labeling models to automatically track paw and facial movement (Abdus-Saboor et al., 2019).

Indeed, the same group from the previous study later introduced two updated pain analysis pipelines that automatically track paw and body movement and then extract and analyze the features of the movement. In the first updated pipeline, SLEAP was used to track paw movement and then various algorithms were used to extract information from the paw movement, including peak paw withdrawal height, and guarding duration (Jones et al., 2020). Then, a univariate pain score is generated using ordinal logistic regression. This first pipeline was validated with the presentation of different

FIGURE 1
Overview of an example workflow for deep learning (DL)-assisted automated pain/withdrawal behavioral analysis. Animal behavior is first recorded on video, including both neurotypical behavior and potential pain/withdrawal behavior. Distinct frames are then sampled from the video pool, and the points of interest in the frames are manually annotated. The labeled frames are then used to train an encoder-decoder convolutional neural network (CNN) with tools such as DeepLabCut or SLEAP. Once the model is trained and achieves a desired level of accuracy, the full videos are fed into the model to generate pose-estimation data for all mice. Finally, behavioral information is extracted from the estimated pose data and quantified ahead of statistical comparisons. This extracted behavioral information can then be fed into field-standard global scoring algorithms and models, thus allowing for comparison of pain/withdrawal behavior between control and experimental groups of mice. Importantly, there are many easily accessible and open-source tools for each step of the example analysis pipeline, many of which have been tested and validated in the context of translational pain and opioid abuse research. Figure created with BioRender.com.

innocuous and noxious stimuli and the chemogenic activation of the pain assembly in the basolateral amygdala (Jones et al., 2020). In the second updated pipeline, the same group uses a wide group of tools and algorithms, including high speed behavioral tracking, time-of-flight infrared tracking, DeepLabCut, principal component analysis, B-SOiD, MoSeq, and learned embedding with doc2vec, to automatically extract and quantify allodynia and other pain behaviors in mice (Bohic et al., 2021).

Another line of work has sought to automate the application of the mouse grimace scale (MGS), which uses a collection of five facial features to characterize the presence of pain (Langford et al., 2010). Because MGS scoring requires trained individuals to provide the analyses, automation of this scoring was well-needed (Kopaczka et al., 2018). In the first example of automated MGS analysis, one group used a U-Net architecture to segment

the regions of interest, including the ears, eyes, and body. The regions of interest are then fed into a CNN that outputs to a single unit with a linear regression classifier to label input images with 0–9 grimace scores (based on the 0–1 regression output) (Kopaczka et al., 2018). Moreover, the authors provide a GUI and server-client architecture to access automated and real-time tracking of behavioral experiments. The model that they developed had a 0.871 mean absolute error on their 0–9 scale (Kopaczka et al., 2018). A second group trained a CNN with MGS labels to identify the presence of pain (Andresen et al., 2020). With accuracy levels approaching >90%, the group was able to accurately identify the presence of postoperative pain in mice (Andresen et al., 2020). Lastly, another group used transfer learning from an InceptionV3 architecture that was pretrained on ImageNet (Tuttle et al., 2018). The final FC and two Softmax layers were retrained to provide "pain"/"no pain" confidence

scores. When analyzing their entire test image set, the model had a reported accuracy of 84%.

Automated analysis of opioid withdrawal behavior

Opioid use disorder is a chronic debilitating condition driven by a cycle of compulsive opioid use, withdrawal, craving, and relapse (Strang et al., 2020). Importantly, opioid dependence and opioid withdrawal syndrome are fundamental features of OUD. These features are thought to induce long-term behavioral and physiological changes as well as increase motivation and propensity for future opioid abuse (Strang et al., 2020). In preclinical research, opioid dependence, and withdrawal are key aspects of modeling OUD. In rodents, opioid dependence is generally achieved through either continual access (e.g., volitional self-administration) or through a paradigm of escalating doses of injections (i.e., non-volitional exposure) (Spanagel, 2022). Opioid dependence in rodents has been shown to produce a range of withdrawal symptoms, including increased grooming, wet-dog shakes, tremors, jumping, rearing, wall climbing, diarrhea, and anxiety-like behavior, among others (Bravo et al., 2020, 2021; Gipson et al., 2020). Along with a withdrawal-like syndrome, opioid dependence in rodents also induces opioid-induced hyperalgesia (Liang et al., 2006; Roeckel et al., 2016).

As with pain behavior scoring, manual scoring of rodent withdrawal behavior involves visual inspection and timing of discrete behavioral events (e.g., grooming, wet-dog shakes, tremors, etc.). Given the drawbacks of manual scoring, automated analysis techniques offer the ability to increase efficiency and decrease error in the assessment of opioid withdrawal in preclinical models of OUD (Fried et al., 2020). However, there has been a remarkable absence of automated analysis in this context. Notably, only one identified study focused on the automated analysis of opioid withdrawal behavior. In this preliminary study, Murphy and colleagues set out to establish a translational model of oral oxycodone self-administration using automated and open-source tools (Murphy et al., 2021). Specifically, markerless pose estimation and subsequent supervised ML predictive classifiers for opioid withdrawal-related behavior were used in C57BL/6J mice that were trained to self-administer oxycodone orally for 12 days with *ad libitum* access. DeepLabCut was used for markerless pose estimation of head, paw, torso, and tail movements. Quantification and extraction of behavioral information were subsequently performed using SimBA, thus allowing Murphy and colleagues to assess jumping, climbing, rearing, grooming, tremors, and Euclidean distance of displacement of all body parts. Ultimately, this automated withdrawal behavioral analysis pipeline was able to detect and quantify physical signs

of dependence that are consistent with opioid withdrawal syndrome in mice (Murphy et al., 2021).

To our knowledge only a single paper has been published that explicitly focused on the automated analysis of opioid withdrawal behaviors; however, other pipelines have been built to track behaviors that are related to opioid withdrawal (Kopaczka et al., 2019). For example, as described, one group built a two-stream model with the ResNet101 architecture to classify general activity patterns in mice, including general movement, grooming, resting, turning, and facing away from the camera (Kopaczka et al., 2019). Importantly, the grooming and movement classifications from this model could be generalized and applied to opioid abuse research. Interestingly, another study automated the analysis of rat ultrasonic vocalizations during a chronic fentanyl self-administration paradigm (Dao et al., 2021). This was achieved using DeepSqueak, a DL-based system for analyzing ultrasonic vocalizations (Coffey et al., 2019). Given that ultrasonic vocalizations are thought to correlate with rodent affect, this tool could be valuable in analyzing another translational aspect of withdrawal syndrome. Finally, other models described in the previous sections, such as the paw flick (Wotton et al., 2020) or scratching (Kobayashi et al., 2021) classification models could also be generalized and used in the context of opioid dependence and withdrawal.

Limitations

As no tool is perfect, it is important to note that these automated pipelines have their own sets of pitfalls, in common with manual scoring. Automated models still require manual labeling of training data (although frankly minimal, e.g., 100–150 frames for accurate DeepLabCut models). The frames that are automatically labeled and automated analysis frameworks must also be manually validated, which means there is still a need for local expertise over the behaviors that are being automatically scored. Moreover, many of these pipelines require a certain degree of technical and computational proficiency and technological specifications. In terms of confirming model tracking and analysis accuracy, an additional limitation is that there is currently no unified database of labeled pain or OUD behavioral recordings that can be used for model validations. Nonetheless, automated behavioral tracking and analysis pipelines are being actively improved, and their associated pitfalls are being minimized. In that regard, we argue that it is important for the developers of automated behavioral analysis tools to consider the end-user in mind when creating and deploying automated analysis packages. Future tools should be open-source and easily accessible in public codebase repositories, and tools should ideally have high-level APIs or even GUIs for accessible deployment.

Future directions

Automated behavioral tracking and analysis tools are well-positioned to blossom in the fields of pain and opioid abuse research. We argue that the accuracy and validation testing of these tools have demonstrated their feasibility in the contexts of pain and OUD research. Particularly in the context of pain research, the articles described in this review have demonstrated the potential to classify and quantify several nuanced behaviors with high levels of accuracy. Preliminary work in the context of OUD has demonstrated similar effectiveness of DL and ML tools for automated scoring of withdrawal-related behaviors in mice. These already highly accurate and primarily open-source automated pipelines discussed here will act as the foundation for a future of automated pain and OUD behavioral testing (**Supplementary Table 1**).

Taken together, ML and DL automated behavioral analysis pipelines have proven to be highly powerful across many fields. With their increased implementation, automated ML and DL-based automated behavioral analysis have the potential to increase the efficiency of pre-clinical investigation of pain and opioid abuse, as well as increase its rigor and reproducibility. Now is the time for scientists working in the fields of pain and OUD research to move beyond the validation stages and begin implementing automated behavioral analysis tools to test novel research questions. In doing so, researchers may be afforded an opportunity to ask more complex and comprehensive questions in their pursuit of novel pain therapeutics with less abuse potential.

Author contributions

JB, DB-K, and RN designed, wrote, and edited the manuscript. RW helped with research and writing of the manuscript. DB-K and JB designed the figure. All authors contributed to the article and approved the submitted version.

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.

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.

Author disclaimer

The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health.

Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fnins.2022.953182/full#supplementary-material

References

Abdus-Saboor, I., Fried, N. T., Lay, M., Burdge, J., Swanson, K., Fischer, R., et al. (2019). Development of a mouse pain scale using sub-second behavioral mapping and statistical modeling. *Cell Rep.* 28, 1623.–1634. doi: 10.1016/j.celrep.2019.07.017

Ahmad, F. B., Rossen, L. M., and Sutton, P. (2021). *Provisional drug overdose death counts. National Center for Health Statistics.* Hyattsville, MD: National Center for Health Statistics.

Anderson, D. J., and Perona, P. (2014). Toward a science of computational ethology. *Neuron* 84, 18–31. doi: 10.1016/j.neuron.2014.09.005

Andresen, N., Wöllhaf, M., Hohlbaum, K., Lewejohann, L., Hellwich, O., Thöne-Reineke, C., et al. (2020). Towards a fully automated surveillance of well-being status in laboratory mice using deep learning: Starting with facial expression analysis. *PLoS One* 15:e0228059. doi: 10.1371/journal.pone.0228059

Baliki, M. N., and Apkarian, A. V. (2015). Nociception, pain, negative moods, and behavior selection. *Neuron* 87, 474–491.

Bohic, M., Pattison, L. A., Jhumka, Z. A., Rossi, H., Thackray, J. K., and Ricci, M. (2021). Mapping the signatures of inflammatory pain and its relief. *bioRxiv* [Preprint]. doi: 10.1101/2021.06.16.448689

Brattoli, B., Büchler, U., Dorkenwald, M., Reiser, P., Filli, L., Helmchen, F., et al. (2021). Unsupervised behaviour analysis and magnification (uBAM) using deep learning. *Nat. Mach. Intell.* 3, 495–506.

Bravo, I. M., Bluitt, M., and McElligott, Z. A. (2021). Examining opioid withdrawal scoring and adaptation of global scoring systems to male and female C57BL/6J mice. *bioRxiv*. doi: 10.1101/2021.10.11.463944

Bravo, I. M., Luster, B. R., Flanigan, M. E., Perez, P. J., Cogan, E. S., Schmidt, K. T., et al. (2020). Divergent behavioral responses in protracted opioid withdrawal

in male and female C57BL/6J mice. *Eur. J. Neurosci.* 51, 742–754. doi: 10.1111/ejn.14580

Coffey, K. R., Marx, R. G., and Neumaier, J. F. (2019). DeepSqueak: a deep learning-based system for detection and analysis of ultrasonic vocalizations. *Neuropsychopharmacology* 44, 859–868. doi: 10.1038/s41386-018-0303-6

Dao, A. N., Beacher, N. J., Mayr, V., Montemarano, A., Hammer, S., and West, M. O. (2021). Chronic fentanyl self-administration generates a shift toward negative affect in rats during drug use. *Brain Sci.* 11:1064. doi: 10.3390/brainsci11081064

Datta, S. R., Anderson, D. J., Branson, K., Perona, P., and Leifer, A. (2019). Computational neuroethology: a call to action. *Neuron* 104, 11–24. doi: 10.1016/j.neuron.2019.09.038

Deuis, J. R., Dvorakova, L. S., and Vetter, I. (2017). Methods used to evaluate pain behaviors in rodents. *Front. Mol. Neurosci.* 10:284. doi: 10.3389/fnmol.2017.00284

Dunn, T. W., Marshall, J. D., Severson, K. S., Aldarondo, D. E., Hildebrand, D. G. C., and Chettih, S. N. (2021). Geometric deep learning enables 3D kinematic profiling across species and environments. *Nat. Methods* 18, 564–573. doi: 10.1038/s41592-021-01106-6

Fried, N. T., Chamessian, A., Zylka, M. J., and Abdus-Saboor, I. (2020). Improving pain assessment in mice and rats with advanced videography and computational approaches. *Pain* 161:1420. doi: 10.1097/j.pain.0000000000001843

Gipson, C. D., Dunn, K. E., Bull, A., Ulangkaya, H., and Hossain, A. (2020). Establishing preclinical withdrawal syndrome symptomatology following heroin self-administration in male and female rats. *Exp. Clin. Psychopharmacol.* 29, 636–649. doi: 10.1037/pha0000375

Hsu, A. I., and Yttri, E. A. (2021). B-SOiD, an open-source unsupervised algorithm for identification and fast prediction of behaviors. *Nat. Comm.* 12:5188. doi: 10.1038/s41467-021-25420-x

Hu, Y., Ferrario, C. R., Ghimire, A., Watson, B., Iwasaki, K., White, H., et al. (2022). LabGym: A versatile computational tool for automatic quantification of user-defined animal behaviors. *bioRxiv* [Preprint]. doi: 10.1101/2022.02.17.480911

Jones, J. M., Foster, W., Twomey, C. R., Burdge, J., Ahmed, O. M., Pereira, T. D., et al. (2020). A machine-vision approach for automated pain measurement at millisecond timescales. *elife* 9:e57258. doi: 10.7554/eLife.57258

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

Kobayashi, K., Matsushita, S., Shimizu, N., Masuko, S., Yamamoto, M., and Murata, T. (2021). Automated detection of mouse scratching behaviour using convolutional recurrent neural network. *Sci. Rep.* 11, 1–10.

Kopaczka, M., Ernst, L., Schock, J., Schneuing, A., Guth, A., Tolba, R., et al. (2018). "Introducing CNN-based mouse grim scale analysis for fully automated image-based assessment of distress in laboratory mice," *Proceedings of the Eurographics Workshop on Visual Computing for Biology and Medicine* EG VCBM* 18, 101–106.

Kopaczka, M., Tillmann, D., Ernst, L., Schock, J., Tolba, R., and Merhof, D. (2019). "Assessment of laboratory mouse activity in video recordings using deep learning methods," in *2019 41st Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, (Piscataway, NJ) 3673–3676.

Langford, D. J., Bailey, A. L., Chanda, M. L., Clarke, S. E., Drummond, T. E., Echols, S., et al. (2010). Coding of facial expressions of pain in the laboratory mouse. *Nat. Methods* 7, 447–449.

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, 1–9. doi: 10.1038/s41592-022-01443-0

Liang, D.-Y., Liao, G., Wang, J., Usuka, J., Guo, Y., Peltz, G., et al. (2006). A genetic analysis of opioid-induced hyperalgesia in mice. *J. Am. Soc. Anesthesiol.* 104, 1054–1062.

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

Mathis, M. W., and Mathis, A. (2020). Deep learning tools for the measurement of animal behavior in neuroscience. *Curr. Opin. Neurobiol.* 60, 1–11.

McInnes, L., Healy, J., and Melville, J. (2018). Umap: Uniform manifold approximation and projection for dimension reduction. *arXiv* [Preprint]. arXiv:1802.03426. doi: 10.1016/j.csbj.2022.06.051

Murphy, C. A., Chang, Y.-H., Pareta, R., Li, J.-N., Earnest, T., Tooley, J., et al. (2021). Modeling features of addiction with an oral oxycodone self-administration paradigm. *bioRxiv* [Preprint]. doi: 10.1101/2021.02.08.430180

Nadeau, S. E., Wu, J. K., and Lawhern, R. A. (2021). Opioids and chronic pain: an analytic review of the clinical evidence. *Front. Pain Res.* 44:721357. doi: 10.3389/fpain.2021.721357

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

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.

Roeckel, L.-A., Le Coz, G.-M., Gavériaux-Ruff, C., and Simonin, F. (2016). Opioid-induced hyperalgesia: cellular and molecular mechanisms. *Neuroscience* 338, 160–182.

Skelly, A. C., Chou, R., Dettori, J. R., Turner, J. A., Friedly, J. L., Rundell, S. D., et al. (2020). *Noninvasive Nonpharmacological Treatment for Chronic Pain: a Systematic Review Update.* Report No.: 20-EHC009. Rockville, MD: Agency for Healthcare Research and Quality (US).

Spanagel, R. (2022). Animal models of addiction. *Dialog. Clin. Neurosci.* 19, 247–258.

Strang, J., Volkow, N. D., Degenhardt, L., Hickman, M., Johnson, K., Koob, G. F., et al. (2020). Opioid use disorder. *Nat. Rev. Dis. Prim.* 6:3.

Sturman, O., von Ziegler, L., Schläppi, C., Akyol, F., Privitera, M., Slominski, D., et al. (2020b). 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

Sturman, O., von Ziegler, L., Schläppi, C., Akyol, F., Grewe, B., and Bohacek, J. (2020a). Deep learning based behavioral analysis enables high precision rodent tracking and is capable of outperforming commercial solutions. *bioRxiv* [Preprint].

Tuttle, A. H., Molinaro, M. J., Jethwa, J. F., Sotocinal, S. G., Prieto, J. C., Styner, M. A., et al. (2018). A deep neural network to assess spontaneous pain from mouse facial expressions. *Mol. Pain* 14:1744806918763658. doi: 10.1177/1744806918763658

Volkow, N. D., and McLellan, A. T. (2016). Opioid abuse in chronic pain—misconceptions and mitigation strategies. *N. Engl. J. Med.* 374, 1253–1263. doi: 10.1056/NEJMra1507771

Wiltschko, A. B., Johnson, M. J., Iurilli, G., Peterson, R. E., Katon, J. M., Pashkovski, S. L., et al. (2015). Mapping sub-second structure in mouse behavior. *Neuron* 88, 1121–1135. doi: 10.1016/j.neuron.2015.11.031

Wiltschko, A. B., Tsukahara, T., Zeine, A., Anyoha, R., Gillis, W. F., Markowitz, J. E., et al. (2020). Revealing the structure of pharmacobehavioral space through motion sequencing. *Nat. Neurosci.* 23, 1433–1443. doi: 10.1038/s41593-020-00706-3

Wotton, J. M., Peterson, E., Anderson, L., Murray, S. A., Braun, R. E., Chesler, E. J., et al. (2020). Machine learning-based automated phenotyping of inflammatory nocifensive behavior in mice. *Mol. Pain* 16:1744806920958596. doi: 10.1177/1744806920958596