

NEWS

Six steps to using machine learning for animal behavior research

BY NIKO MCCARTY

26 MAY 2022

Listen to this story:

<https://www.spectrumnews.org/wp-content/uploads/2022/05/audio-bc2df1b8-52d8-44ae-968c-276d71b31ea1-encodings.mp3>

Just a few years ago, **Nastacia Goodwin** spent most days sitting at a computer in a lab at Smith College in Northampton, Massachusetts, stopwatch in hand, eyes fixed on three-hour long videos of prairie voles. Whenever an animal huddled close to another — *click* — she recorded the duration of their interaction.

It didn't take long before Goodwin, now a graduate student in **Sam Golden's research group** at the University of Washington in Seattle, became eager to find a faster, less biased way to annotate videos. Machine learning was a logical choice.

Goodwin co-developed **Simple Behavioral Analysis**, or SimBA, an **open-source tool** to automatically detect and classify animal behaviors from videos. It's one of a growing number of machine-learning techniques that are accelerating behavioral neuroscience and making results more reproducible.

"Now I can score several hundred of those [videos] each week, just by clicking a few buttons," Goodwin says.

Such tools have already been used to **quantify marble-burying 'bouts'** — a proxy for **repetitive behaviors** — in autism mouse models, and to track mice with **mutations in the gene CNTNAP2** during the open field test, a gauge for anxiety-like behaviors.

Goodwin and other experts describe three basic steps for using machine-learning tools to analyze videos of animal behavior. First, upload the videos to a pose-estimation tool, such as **Social LEAP**

Estimates Animal Poses (SLEAP) or **DeepLabCut**, both of which are animal-agnostic. These tools pinpoint each body part of an animal — the tail, head and limbs — for each video frame and then output the data as x and y coordinates.

Next, upload those coordinates to a behavioral classifier, such as SimBA or the **Janelia Automatic Animal Behavior Annotator** (JAABA). These tools “look at the relationship between those points and assess how it changes over time,” Goodwin says. With guidance from a researcher, these classification algorithms ‘learn’ to identify specific behaviors based on the coordinates.

In the final step, researchers refine the behavioral classification algorithms by slightly tweaking the parameters to improve accuracy.

Spectrum asked researchers for advice on using machine learning to quantify animal behaviors.

1. Start small and cheap.

High-resolution cameras are usually overkill and not required to automate video analyses, says Sam Golden, assistant professor of neuroscience at the University of Washington. Start with a simple webcam instead.

“What you’re really looking for is the minimum viable product in terms of resolution and frame rate,” he says. “You don’t need really high-end camera hardware, or even computational hardware anymore, to dive in and do it.”

Starting cheap shouldn’t come at the cost of quality, though. Videos should be **carefully collected**. “Reflections can be a real problem for many algorithms if there’s light scatter or shadowing,” says **Mark Laubach**, professor of neuroscience at American University in Washington, D.C. “The deep learning may not be learning from the animal; it may be learning from reflections in the box.”

2. You don’t need to know how to code. But it helps.

Most tools, including SLEAP, DeepLabCut and SimBA, don’t require any programming. Each offers stand-alone software with easy-to-use interfaces.

“We front-loaded a lot of engineering work to make all this stuff really, really smooth,” says **Talmo Pereira**, Salk fellow at the Salk Institute for Biological Studies in La Jolla, California, and co-creator of SLEAP. “I think we’ve made it extremely easy for you to be able to spend, like, less than an afternoon and start getting real results.”

The tools work for almost any organism, too. Teams have used SLEAP with mice, flies, whales and bacteria, Pereira says.

Researchers with programming skills can do more with their data. Assessing an algorithm's accuracy, for instance, often requires bespoke code and at least a basic understanding of machine-learning models and statistics. "I'm able to produce quality research and data so much faster because I can code," Goodwin says.

3. Apply the tools wisely.

Machine learning isn't always useful. Golden points to the **elevated plus maze**, used to measure anxiety, as a prime example. "We don't need a behavioral classifier for when an animal is transitioning into an open and closed space," he says. "We can just track the animal and look at when it's in a region of interest."

And if a behavior is difficult to detect by eye, a machine-learning algorithm probably won't be able to detect it either, Golden says. "If you have to squint when you're looking at a behavior, or use your human intuition to sort of fill in the blanks, you're not going to be able to generate an accurate classifier from those videos."

Machine learning excels at accurately and quickly detecting and quantifying easy-to-spot behaviors, such as grooming and social interactions. "It's like going from the original versions of DNA sequencing to the next-generation technologies. It lets you operate at a scale that wasn't feasible before," says **Ann Kennedy**, assistant professor of neuroscience at Northwestern University in Chicago, Illinois. "You can hire people to annotate videos, but people aren't consistent. Their attention drifts over time."

In a *Nature Neuroscience* paper from 2020, researchers gave nearly **700 different mice** one of 16 drugs. Using a **tool called MoSeq**, which classifies behaviors much like JAABA and SimBA do, the researchers "were able to distinguish which drug in which concentration each animal was given, just based on changes in behavior expression," Kennedy says.

4. Test, test, test your algorithm.

Some behavioral classification algorithms, such as SimBA, use supervised machine-learning techniques, meaning that they 'learn' from curated datasets. How a researcher assembles those datasets determines how accurate these algorithms will be. To use these algorithms, researchers create three datasets: training, validation and testing.

The training dataset, as advertised, trains the initial machine learning models. These models return time stamps wherever they 'think' that a specified behavior is occurring in a video. Sometimes these time stamps aren't quite right on the first pass.

Researchers then use the validation dataset to refine and tweak the model's parameters. Finally, they benchmark the model against a completely new set of videos — the testing dataset — to

quantify its accuracy.

The best way to improve accuracy, according to Kennedy, is to create the training dataset using videos from many different animals, and to use video frames that are far apart from each other in time.

“Neighboring video frames are super correlated with each other. If your test set is 10 percent of your data, then most of the time, any given frame in your test set is going to have a neighboring frame in your training set,” Kennedy says. “They’re not true independent observations. And that can lead to overfitting of the training data and really bad over-reporting of classifier performance.”

It’s easy to assess the accuracy of pose-estimation coordinates, says **Mackenzie Mathis**, assistant professor of neuroscience at the Swiss Federal Institute of Technology in Lausanne, Switzerland, and co-creator of DeepLabCut, because the positions of body parts are based on a “ground truth” — researchers can see, with their eyes, whether they are accurately labeled.

Assessing the accuracy of a behavioral classifier, though, is far more difficult, because different behaviors can look similar, Mathis says: “Where it gets tricky is if someone tests an algorithm to detect whether a mouse is freezing because it’s afraid, or that mouse is just standing there because it’s bored, or that mouse is asleep.”

5. Explainability matters.

Two research groups that use machine-learning tools to classify the same behavior will, in many cases, produce two algorithms that detect the same behavior based on completely different data patterns.

In one instance, Goodwin trained a classifier to detect when two mice attack each other. “But then I started filming some attacks in a slightly different arena. And suddenly, my animals, instead of attacking from the side, were running around side by side in the cage.” The algorithm classified that side-by-side running as an attack because the mice “were running fast, and they’re close together,” even though it clearly wasn’t an attack.

“Understanding that my classifiers were really just focusing on proximity and speed,” Goodwin says, “showed me that this will not generalize to new settings.” By adding a few dozen video frames of the mice running side by side, and labeling them as ‘not an attack,’ the classifier’s accuracy improved.

Explainability metrics are quantitative tools that describe exactly how some feature — such as the proximity of one mouse’s nose to the tail of another — contributes to an algorithm’s decision-making process. When conveying an algorithm’s findings, researchers should incorporate these metrics, because they make for more transparent results.

6. Ask for help. Be patient.

Thousands of researchers use machine-learning tools to quantify animal behaviors. DeepLabCut alone has been downloaded more than **365,000 times**, according to GitHub, suggesting that there is a sizeable community of users to help troubleshoot issues. New users can turn to a Slack group called **NeuroMethods** to ask for help. SimBA also has an **online forum**, with more than 170 members, where researchers answer questions and share resources.

Laubach, separately, maintains a website called **OpenBehavior**, a repository of “open-source tools for advancing behavioral neuroscience.” The website also hosts a collection of animal videos to help train machine-learning algorithms.

Funding and workshops, Golden says, are also available through the U.S. National Institutes of Health and **The Jackson Laboratory** (JAX), a nonprofit repository of mouse strains in Bar Harbor, Maine. JAX plans to hold a **four-day course** in October on using machine learning for automated behavioral quantification.

Even with help, it takes patience to master machine-learning basics. Laubach has had three masters students so far who each wanted to use machine learning for their thesis. One student “totally failed,” one student “did OK,” and “one is now using it for his capstone project,” he says. “Learning takes time.”

Cite this article: <https://doi.org/10.53053/DTVA2673>