PAIN

Painless pain assessments with machine learning

Tuttle et al. *Molecular Pain* 14, 1–9, 2018.

The grimace scales, first published for the mouse in 2010 by Jeff Mogil's lab at McGill University (*Nat Methods* 7, 447–449; 2010), were developed as a way to quantify pain in research animals. The idea has since been adapted for about a dozen different species, lab and otherwise. The scales have been a useful tool for veterinarians and technicians monitoring pain in their charges, but also for basic and preclinical researchers like Mark Zylka at the University of North Carolina (UNC) who are interested in understanding the properties of pain and how to treat it. Like many repetitive tasks, the potential was always there for automation. Zylka first heard of the mouse grimace scale from a presentation Mogil gave before the first paper was officially released, but shelved his own early thoughts about automating it—the time just wasn't right yet.

But as computer algorithms have become increasingly adept at identifying human faces in recent years, the idea returned. If machines could distinguish the features of a human, why not those of a mouse?

Zylka recruited Alex Tuttle, a veteran of Mogil's lab and an experienced pain-scale scorer, as a postdoctoral fellow and charged him with the task of automating the mouse grimace scale. A particularly tech-savvy undergraduate at UNC at the time, Mark Molinaro, soon joined in. Two years of development later, using Google's InceptionV3 neural net and with the help of thousands of training images from Mogil's lab (the very same used to create the original scale) and several thousand more generated in Zylka's lab in North Carolina, the team presents Version 1.0 of an automated mouse grimace scale in the journal *Molecular Pain*.

"What this has effectively done," says Zylka, "is taken what was a very low throughput, tedious assay that needed to be done by highly trained humans and made it possibly the fastest assay you can do in the pain field that's incredibly objective."

Manually, the number of images a well-trained human, who will have spent several weeks learning the signs of the scale, could review and score from recordings was relatively small. "About one image for every

three minutes for a human to score, so you would have about 10 to 15 images per half an hour dataset," Tuttle estimates. "We were able to bump that up to one image per every five seconds, or two seconds, whatever you wanted to do. We set the parameters to have it score however many images we wanted within the timeframe."

Once the team felt confident in the computer's confidence at scoring pain, they tested its predictive value. To do so, the machine learning algorithm was run against novel images of mice that underwent a (real or sham control) laparotomy, with and without post-operative analgesia, to see if it could pick up pain and the alleviation of it in different animals. And it could—mice that underwent the actual surgery showed more signs of pain than the sham group, and pain was relieved with carprofen (though mice falling asleep was a complication that required the team to shorten the recording duration—the analgesic seemed less effective in a 30-minute window following the procedure).

In its current iteration, the algorithm does have a few initial limitations. Not unlike human scorers, it had difficulty distinguishing pain in intermediate ranges, so it is currently tuned to make obvious pain/no pain assessments; larger training sets will be needed to increase that granularity.

It's also limited to a single species with a single coat color at the moment. Mogil prefers outbred mice for his own research, so white-coated CD1 mice made up the initial training sets. Tuttle and Zylka hope that even deeper learning neural networks that are currently being developed by computer scientists may hold the key to transferring the algorithm from coat-to-coat, and species-to-species, without requiring such large initial datasets.

In addition to expanding what the automated Mouse Grimace Scale will eventually be able to handle, the team wants to make sure it's reproducible. "The million dollar question is whether the current version is practical for multiple different labs across the world to use," says Tuttle. Though

How's he feeling? Automation may help make the call in the future. Credit: PetlinDmitry/ iStock / Getty Images Plus

white lab mice may not be as prevalent in labs as say C57BL/6 strains, the team does have some beta testers ready to try out the algorithm.

Those include Mogil, whose own lab is currently looking to install and learn the software components needed to run the automated scoring. Though cautious that there's still room for improvement, Mogil thinks it'll be a useful tool that may help increase uptake of the grimace scale, particularly among those who may perceive it as too complicated or time-consuming to do manually. "If the development of a completely automated method for doing scoring becomes available and widely adopted, I think people will jump on board, as long as they trust it," he says.

In all though, "it's another example of a clever use for [artificial intelligence]," Mogil says, like Zylka noting how machine learning is, after decades of predictions, starting to take off. "It's just a matter of people coming up with new applications and this was a pretty good one."

Recognizing pain will always be an important consideration for anyone working with animals in the lab. The future may help make identifying it just a little less of a pain.

Ellen P. Neff

Published online: 23 May 2018
<https://doi.org/10.1038/s41684-018-0078-8>