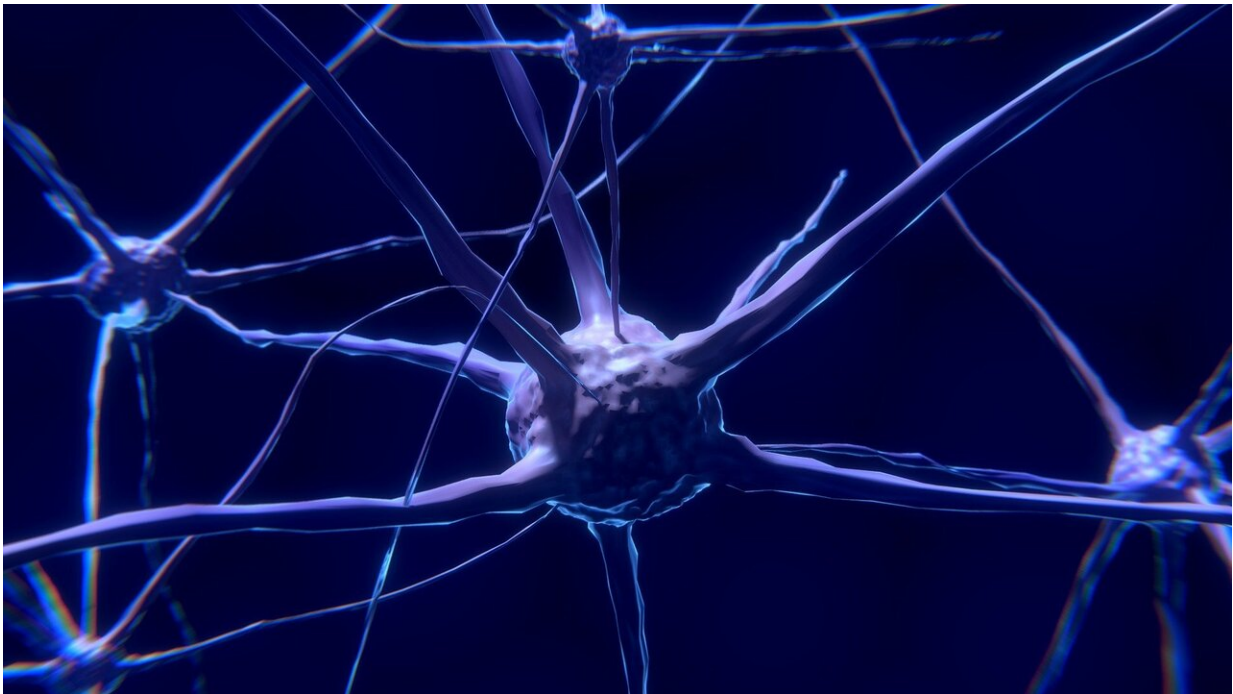# Breaking stereotype: Brain models are not one-size-fits-all

August 25 2022

Credit: Pixabay/CC0 Public Domain

Machine learning has helped scientists understand how the brain gives rise to complex human characteristics, uncovering patterns of brain activity that are related to behaviors like working memory, traits like impulsivity, and disorders like depression. And with these tools, scientists can create models of these relationships that can then be used, in theory, to make predictions about the behavior and health of

individuals.

But that only works if models represent everyone, and previous research has shown that they don't; for any model, there are some people that the model just doesn't fit.

In a study published Aug. 24 in *Nature*, Yale researchers examined who these models tend to fail, why that happens, and what can be done about it.

For models to be maximally useful, they need to apply to any given individual, says Abigail Greene, an M.D.-Ph.D. student at Yale School of Medicine and lead author of the study.

"If we want to move this kind of work into a clinical application, for example, we need to make sure the model applies to the patient sitting in front of us," she said.

Greene and her colleagues are interested in how models might provide more precise psychiatric characterization, which they think could be achieved in two ways. The first is by better categorizing patient populations. A diagnosis of schizophrenia, for example, encompasses an array of symptoms, and it can look very different from person to person. A deeper understanding of the neural underpinnings of schizophrenia, including its symptoms and subcategories, could allow researchers to group patients in a more nuanced way.

Secondly, there are traits like impulsivity that are shared across a variety of diagnoses. Understanding the neural basis of impulsivity could help clinicians target that symptom more effectively regardless of the disease diagnosis to which it's attached.

"And both advances would have implications for treatment responses,"

said Greene. "The better we can understand these subgroups of individuals who may or may not carry the same diagnoses, the better we can tailor treatments to them."

But first, models need to be generalizable to everybody, she said.

To understand model failure, Greene and her colleagues first trained models that could use patterns of brain activity to predict how well a person would score on a variety of cognitive tests. When tested, the models correctly predicted how well most individuals would score. But for some people, they were incorrect, wrongly predicting people would score poorly when they actually scored well, and vice versa.

The research team then looked at who the models failed to categorize correctly.

"We found that there was consistency—the same individuals were getting misclassified across tasks and across analyses," said Greene. "And the people misclassified in one dataset had something in common with those misclassified in another dataset. So there really was something meaningful about being misclassified."

Next, they looked to see if these similar misclassifications could be explained by differences in those individuals' brains. But there were no consistent differences. Instead, they found misclassifications were related to sociodemographic factors like age and education and clinical factors like symptom severity.

Ultimately, they concluded that the models weren't reflecting the cognitive ability alone. They were instead reflecting more complex "profiles"—sort of mashups of the cognitive abilities and various sociodemographic and clinical factors, explained Greene.

"And the models failed anyone who didn't fit that stereotypical profile," she said.

As one example, models used in the study associated more education with higher scores on the cognitive tests. Any individuals with less education who scored well didn't fit the model's profile and were therefore often erroneously predicted to be low scorers.

Adding to the complexity of the problem, the model did not have access to sociodemographic information.

"The sociodemographic variables are embedded in the cognitive test score," explained Greene. Essentially, biases in how cognitive tests are designed, administered, scored, and interpreted can seep into the results that are obtained. And bias is an issue in other fields as well; research has uncovered how input data bias affects models used in criminal justice and health care, for instance.

"So the test scores themselves are composites of the cognitive ability and these other factors, and the model is predicting the composite," said Greene. That means researchers need to think more carefully about what is really being measured by a given test and, therefore, what a model is predicting.

The study authors provide several recommendations for how to mitigate the problem. At the study design phase, they suggest, scientists should employ strategies that minimize bias and maximize the validity of the measurements they're using. And after researchers collect data, they should as often as possible use statistical approaches that correct for the stereotypical profiles that remain.

Taking these measures will lead to models that better reflect the cognitive construct under study, the researchers say. But they note that

fully eliminating bias is unlikely, so it should be acknowledged when interpreting model output. Additionally, for some measures, it may turn out that more than one model is necessary.

"There's going to be a point where you just need different models for different groups of people," said Todd Constable, professor of radiology and biomedical imaging at Yale School of Medicine and senior author of the study. "One model is not going to fit everybody."

More information: Abigail S. Greene et al, Brain–phenotype models fail for individuals who defy sample stereotypes, *Nature* (2022). DOI: 10.1038/s41586-022-05118-w

Provided by Yale University