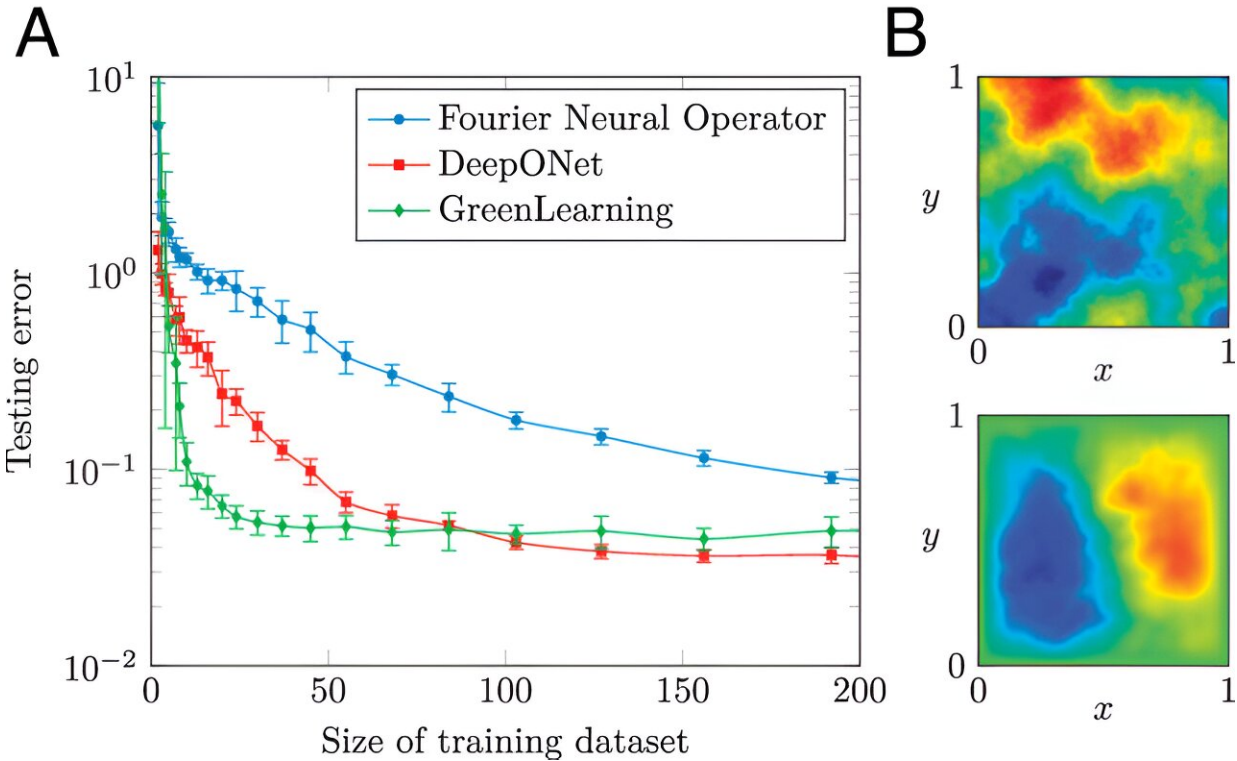


Machine learning models can produce reliable results even with limited training data

September 19 2023, by Sarah Collins



Elliptic PDE learning methods can be data-efficient. Credit: *Proceedings of the National Academy of Sciences* (2023). DOI: 10.1073/pnas.2303904120

Researchers have determined how to build reliable machine learning models that can understand complex equations in real-world situations

while using far less training data than is normally expected.

The researchers, from the University of Cambridge and Cornell University, found that for [partial differential equations](#)—a class of physics equations that describe how things in the [natural world](#) evolve in space and time—machine learning models can produce reliable results even when they are provided with limited data.

Their results, [reported in the *Proceedings of the National Academy of Sciences*](#), could be useful for constructing more time- and cost-efficient machine learning models for applications such as engineering and climate modeling.

Most machine learning models require large amounts of training data before they can begin returning accurate results. Traditionally, a human will annotate a large volume of data—such as a set of images, for example—to train the model.

"Using humans to train machine learning models is effective, but it's also time-consuming and expensive," said first author Dr. Nicolas Boullé, from the Isaac Newton Institute for Mathematical Sciences. "We're interested to know exactly how little data we actually need to train these models and still get reliable results."

Other researchers have been able to train machine learning models with a small amount of data and get excellent results, but how this was achieved has not been well-explained. For their study, Boullé and his co-authors, Diana Halikias and Alex Townsend from Cornell University, focused on partial differential equations (PDEs).

"PDEs are like the building blocks of physics: they can help explain the physical laws of nature, such as how the steady state is held in a melting block of ice," said Boullé, who is an INI-Simons Foundation

Postdoctoral Fellow. "Since they are relatively simple models, we might be able to use them to make some generalizations about why these AI techniques have been so successful in physics."

The researchers found that PDEs that model diffusion have a structure that is useful for designing AI models. "Using a simple model, you might be able to enforce some of the physics that you already know into the training data set to get better accuracy and performance," said Boullé.

The researchers constructed an [efficient algorithm](#) for predicting the solutions of PDEs under different conditions by exploiting the short and long-range interactions happening. This allowed them to build some mathematical guarantees into the model and determine exactly how much training data was required to end up with a robust model.

"It depends on the field, but for physics, we found that you can actually do a lot with a very limited amount of data," said Boullé. "It's surprising how little data you need to end up with a reliable [model](#). Thanks to the mathematics of these equations, we can exploit their structure to make the models more efficient."

The researchers say that their techniques will allow data scientists to open the 'black box' of many machine learning models and design new ones that can be interpreted by humans, although future research is still needed.

"We need to make sure that models are learning the right things, but machine learning for physics is an exciting field—there are lots of interesting math[s] and [physics](#) questions that AI can help us answer," said Boullé.

More information: Nicolas Boullé et al, Elliptic PDE learning is provably data-efficient, *Proceedings of the National Academy of Sciences*

(2023). [DOI: 10.1073/pnas.2303904120](https://doi.org/10.1073/pnas.2303904120)

Provided by University of Cambridge

Citation: Machine learning models can produce reliable results even with limited training data (2023, September 19) retrieved 27 April 2024 from <https://techxplore.com/news/2023-09-machine-reliable-results-limited.html>

This document is subject to copyright. Apart from any fair dealing for the purpose of private study or research, no part may be reproduced without the written permission. The content is provided for information purposes only.