



Machine Learning Used to Predict Synthesis of Complex Novel Materials

2022-01-02

Scientists and institutions dedicate more resources each year to the discovery of novel materials to fuel the world. As natural resources diminish and the demand for higher value and advanced performance products grows, researchers have increasingly looked to nanomaterials.

Nanoparticles have already found their way into applications ranging from energy storage and conversion to quantum computing and therapeutics. But given the vast compositional and structural tunability nanochemistry enables, serial experimental approaches to identify new materials impose insurmountable limits on discovery.

Now, researchers at [Northwestern University](#) and the Toyota Research Institute ([TRI](#)) have successfully applied machine learning to guide the synthesis of new nanomaterials, eliminating barriers associated with materials discovery. The highly trained algorithm combed through a defined dataset to accurately predict new structures that could fuel processes in clean energy, chemical and automotive industries.

“We asked the model to tell us what mixtures of up to seven elements would make something that hasn’t been made before,” said Chad Mirkin, a Northwestern nanotechnology expert and the paper’s corresponding author. “The machine predicted 19 possibilities, and, after testing each experimentally, we found 18 of the predictions were correct.”

The study, “Machine learning-accelerated design and synthesis of polyelemental heterostructures,” was published today (December 22) in the journal Science Advances.

Mirkin is the George B. Rathmann Professor of Chemistry in the Weinberg College of Arts and Sciences; a professor of chemical and biological engineering, biomedical engineering, and materials science and engineering at the McCormick School of Engineering; and a professor of medicine at the Feinberg School of Medicine. He also is the founding director of the [International Institute for Nanotechnology](#).

Mapping the materials genome

According to Mirkin, what makes this so important is the access to unprecedentedly large, quality datasets because machine learning models and AI algorithms can only be as good as the data used to train them.

The data-generation tool, called a “Megalibrary,” was invented by Mirkin and dramatically expands a researcher’s field of vision. Each Megalibrary houses millions or even billions of nanostructures, each with a slightly distinct shape, structure and composition, all positionally encoded on a two-by-two square centimeter chip. To date, each chip contains more new inorganic materials than have ever been collected and categorized by scientists.

Mirkin’s team developed the Megalibraries by using a technique (also invented by Mirkin) called polymer pen lithography, a massively parallel nanolithography tool that enables the site-specific deposition of hundreds of thousands of features each second.

When mapping the human genome, scientists were tasked with identifying combinations of four bases. But the loosely synonymous “materials genome” includes nanoparticle combinations of any of the usable 118 elements in the periodic table, as well as parameters of shape, size, phase morphology, crystal structure and more. Building smaller subsets of nanoparticles in the form of Megalibraries will bring researchers closer to completing a full map of a materials genome.

Mirkin said that even with something similar to a “genome” of materials, identifying how to use or label them requires different tools.

“Even if we can make materials faster than anybody on earth, that’s still a droplet of water in the ocean of possibility,” Mirkin said. “We want to define and mine the materials genome, and the way we’re doing that is through artificial intelligence.”

Machine learning applications are ideally suited to tackle the complexity of defining and mining the materials genome, but are gated by the ability to create datasets to train algorithms in the space. Mirkin said the combination of Megalibraries with machine learning may finally eradicate that problem, leading to an understanding of what parameters drive certain materials properties.

‘Materials no chemist could predict’

If Megalibraries provide a map, machine learning provides the legend.

Using Megalibraries as a source of high-quality and large-scale materials data for training artificial intelligence (AI) algorithms, enables researchers to move away from the “keen chemical intuition” and serial experimentation typically accompanying the materials discovery process, according to Mirkin.

“Northwestern had the synthesis capabilities and the state-of-the-art characterization capabilities to determine the structures of the materials we generate,” Mirkin said. “We worked with TRI’s AI team to create data inputs for the AI algorithms that ultimately made these predictions about materials no chemist could predict.”

In the study, the team compiled previously generated Megalibrary structural data consisting of nanoparticles with complex compositions, structures, sizes and morphologies. They used this data to train the model and asked it to predict compositions of four, five and six elements that would result in a certain structural feature. In 19 predictions, the machine learning model predicted new materials correctly 18 times — an approximately 95% accuracy rate.

With little knowledge of chemistry or physics, using only the training data, the model was able to accurately predict complicated structures that have never existed on earth.

“As these data suggest, the application of machine learning, combined with Megalibrary technology, may be the path to finally defining the materials genome,” said Joseph Montoya, senior research scientist at TRI.

Metal nanoparticles show promise for catalyzing industrially critical reactions such as hydrogen evolution, carbon dioxide (CO₂) reduction and oxygen reduction and evolution. The model was trained on a large Northwestern-built dataset to look for multi-metallic nanoparticles with set parameters around phase, size, dimension and other structural features that change the properties and function of nanoparticles.

The Megalibrary technology may also drive discoveries across many areas critical to the future, including plastic upcycling, solar cells, superconductors and qubits.

A tool that works better over time

Before the advent of megalibraries, machine learning tools were trained on incomplete datasets collected by different people at different times, limiting their predicting power and generalizability. Megalibraries allow machine learning tools to do what they do best — learn and get smarter over time. Mirkin said their model will only get better at predicting correct materials as it is fed more high-quality data collected under controlled conditions.

“Creating this AI capability is about being able to predict the materials required for any application,” Montoya said. “The more data we have, the greater predictive capability we have. When you begin to train AI, you start by localizing it on one dataset, and, as it learns, you keep adding more and more data — it’s like taking a kid and going from kindergarten to their Ph.D. The combined experience and knowledge ultimately dictates how far they can go.”

The team is now using the approach to find catalysts critical to fueling processes in clean energy, automotive and chemical industries. Identifying new green catalysts will enable the conversion of waste products and plentiful feedstocks to useful matter, hydrogen generation, carbon dioxide utilization and the development of fuel cells. Producing catalysts also could be used to replace expensive and rare materials like iridium, the metal used to generate green hydrogen and CO₂ reduction products.

Read the [original article](#) on Northwestern University.