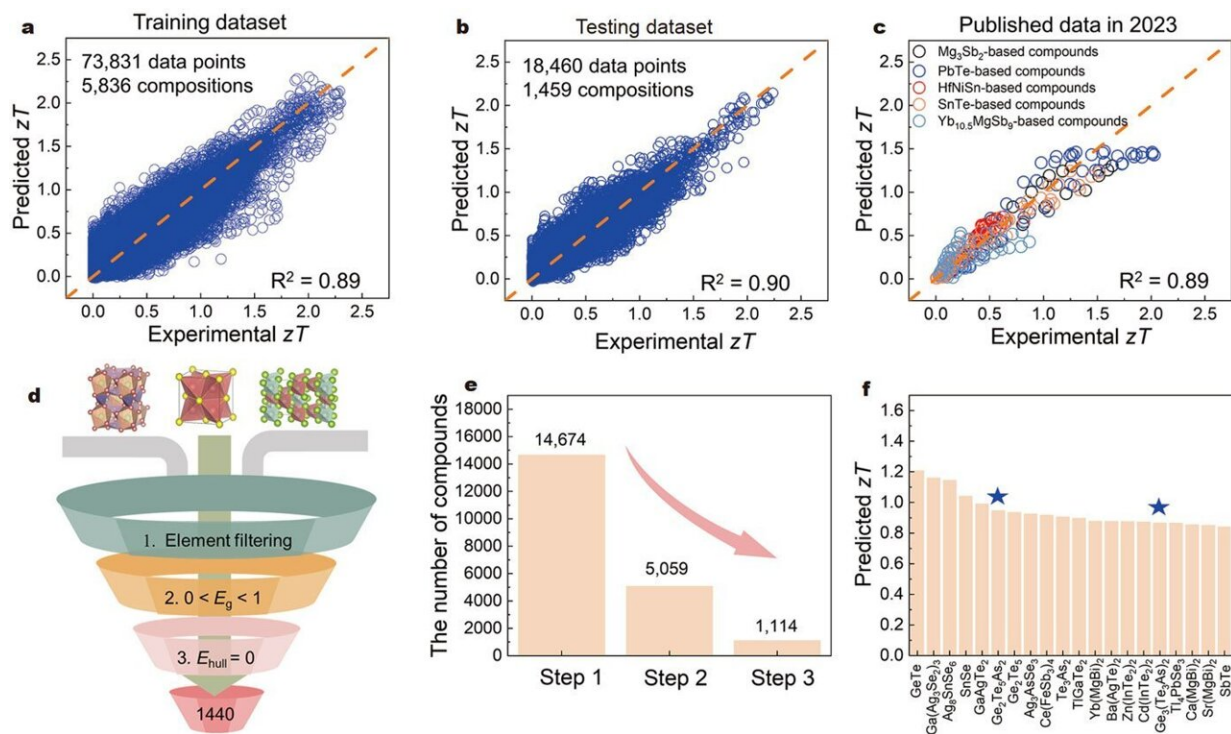


# Rectifying AI's usage in the quest for thermoelectric materials

March 25 2024



Evaluation of ML models and their predictions for new materials. (a–c) Performance of ML model in (a) training dataset, (b) test dataset, and (c) the additional data published in the year of 2023; (d) screening process of selecting materials from the Materials Project dataset; (e) the number of compounds after each screening step; (f) predicted  $zT$  values for the new materials. Credit: *Science China Materials* (2024). DOI: 10.1007/s40843-023-2777-2

A team of researchers has used AI to identify a thermoelectric material with favorable values. The group was able to navigate AI's conventional pitfalls and big data challenges, offering a prime example of how AI can revolutionize materials science. Details were [published](#) in the journal *Science China Materials* on March 8, 2024.

"Traditional methods of finding suitable materials involve trial-and-error, which is time-consuming and often expensive," proclaims Hao Li, associate professor at Tohoku University's Advanced Institute for Materials Research (WPI-AIMR) and corresponding author of the paper. "AI transforms this by combing through databases to identify potential materials that can then be experimentally verified."

Still, challenges remain. Large-scale material datasets sometimes contain errors, and overfitting the predicted temperature-dependent properties is also a common error. Overfitting occurs when a model learns to capture noise or random fluctuations in the [training data](#) rather than the underlying pattern or relationship.

As a result, the model performs well on the training data but fails to generalize new, unseen data. When predicting temperature-dependent properties, overfitting could lead to inaccurate predictions when the model encounters new conditions outside the range of the training data.

Li and his colleagues sought to overcome this to develop a [thermoelectric material](#). These materials convert heat energy into electrical energy, or vice versa. Thus, getting a highly accurate temperature-dependence is critical.

"First, we performed a series of rational actions to identify and discard questionable data, obtaining 92,291 data points comprising 7,295 compositions and different temperatures from the Starrydata2 database—an online database that collects digital data from published papers," states Li.

Then the researchers built machine building models using the Gradient Boosting Decision Tree method. The model achieved remarkable R2 values 0.89, ~0.90, and ~0.89 on the training dataset, test dataset, and new out-of-sample [experimental data](#) released in 2023, demonstrating the models accuracy in predicting newly available materials.

"We could use this model to carry out a large-scale evaluation of the stable materials from the Materials Project database, predicting the potential thermoelectric performance of new materials and providing guidance for experiments," states Xue Jia, Assistant Professor at WPI-AIMR, and co-author of the paper.

Ultimately, the study illustrates the importance of following rigorous guidelines when it comes to data preprocessing and data splitting in [machine learning](#) so that it addresses the pressing issues in [materials science](#). The researchers are optimistic that their strategy can also be applied to other materials, such as electrocatalysts and batteries.

**More information:** Xue Jia et al, Dealing with the big data challenges in AI for thermoelectric materials, *Science China Materials* (2024). [DOI: 10.1007/s40843-023-2777-2](https://doi.org/10.1007/s40843-023-2777-2)

Provided by Tohoku University

Citation: Rectifying AI's usage in the quest for thermoelectric materials (2024, March 25)

retrieved 27 April 2024 from

<https://phys.org/news/2024-03-rectifying-ai-usage-quest-thermoelectric.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.