Leveraging Machine Learning to Navigate Complex Design Spaces in Battery Material Development

Seon-Hwa Lee[1], Insoo Ye[2], Inchul Park[1]
[1] LIB Materials Lab, POSCO NEXT Hub, South Korea

Seonalee110@posco-inc.com

The performance of battery materials is determined by complex interactions between design and process variables. The sheer number of these variables exposes the limitations of traditional one-factor-at-a-time (OFAT) methods, making comprehensive exploration impossible. [1] As a result, industrial optimization relies heavily on domain knowledge, with experts performing transfer learning from the design of similar materials. However, this heuristic approach is limited by human cognitive capabilities and may fail to efficiently navigate the complex design space of the modern battery industry, where controllable dimensions have increased due to rapid advances.

At the same time, economic constraints, such as raw material costs and recycling recoveries, and technical requirements, such as battery performance guarantees, have tightened the constraints on the exploration space. Performance goals have become more stringent in multiple dimensions, requiring greater precision in the optimization process. In this context, the integration of machine learning (ML) and design of experiment (DoE) techniques has been reported to be more effective than traditional Edisonian approaches. [2] ML offers the potential to handle high-dimensional data and discover complex patterns that are not easily identified by empirical methods. [3]

However, issues such as data scarcity and bias due to lack of failure data and anthropogenic biases in existing datasets pose challenges to the effective application of ML in battery material optimization. [4] It is also uncertain whether existing data can help in the design of new materials, as the development of new materials involves predictions that are outside the domain of applicability. Therefore, this study investigates whether ML using transfer learning and active learning can overcome data scarcity and bias to effectively replace the traditional Edisonian approach in the field of battery engineering, especially in the material development stage. [5]

We collected a dataset of design parameters, process variables, and performance data from materials developed over the past several years. The data were collected from a single source using a consistent protocol, which played an important role in reducing data bias. However, our analysis showed that experiments based on domain knowledge inevitably contain anthropogenic biases. The OFAT experiment confirmed data bias, there was a lack of negative data due to psychological factors, and there was also insufficient data on the material we wanted to optimize.

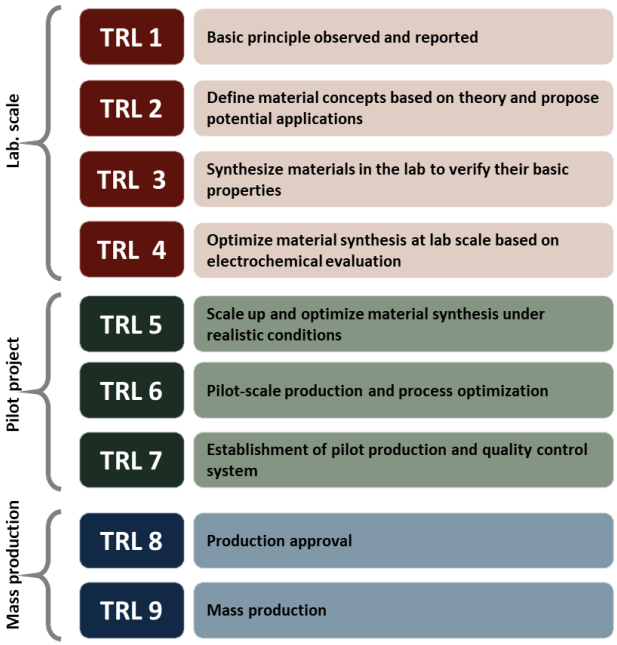

Figure 1. Technical Readiness of Battery Materials

Zero-shot ML models trained on existing materials identified key variables within a narrow chemical space and back-predicted optimal conditions, but they were of little interest because these materials were no longer our focus: the design and processing of new materials outside the domain of applicability are of paramount importance. We used a variety of machine learning algorithms, including neural networks, regression models, and ensemble methods, to design new materials and processes. Despite biased training data and a lack of similar data, we predicted promising optimal conditions. However, identifying key variables to understand the underlying physics remained a challenge.

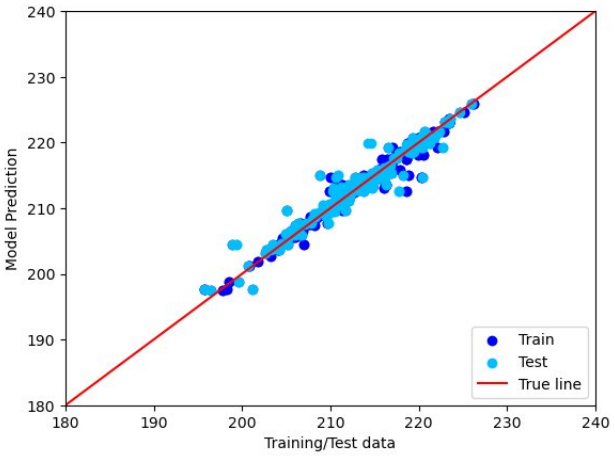

Figure 2. Performance Prediction of Cathode Material for LIB

To overcome this challenge and facilitate efficient transfer learning, we introduced active learning. We ran experiments based on the conditions proposed by the ML model, and then fed the results of these experiments back into the model as additional training data. Through this iterative process, the model made more sophisticated predictions and improved its ability to explore the underlying physics. Despite the initial lack of data, the model's performance improved with each iteration, demonstrating the effectiveness of combining ML and active learning in this context.

The active learning approach helped mitigate the effects of data bias by focusing on the most informative data points, including those underrepresented in the original data set. This led to a better understanding of how certain variables affect performance. Compared to traditional empirical methods, the ML model combined with active learning significantly reduced the number of experiments required to achieve optimal performance. The iterative process streamlined optimization, saving both time and resources. These efficiencies translate into lower costs and accelerated development timelines, critical factors in the rapidly evolving battery industry.

We have shown that ML models combined with active learning can effectively replace traditional Edisonian methods. This has implications for both materials development and the field of data informatics. By accelerating the optimization process, this circular approach motivates the materials development community to collect data, which helps build more robust and comprehensive datasets. This will help strengthen the circular approach to data collection and use, laying the foundation for innovative and efficient materials development strategies.

References

[1] P. Raccuglia *et al.*, Nature, 533 (2016) 73–77
[2] B. Cao *et al.*, ACS Nano, 12, 8 (2018) 7434-7444
[3] T. Xie and JC Grossman, Phys. Rev. Lett., 120 (2018) 145301
[4] X Jia *et al.*, Nature, 573 (2019) 251-255
[5] J Schmidt et al., npj Comput. Mater., 5 (2019) 83