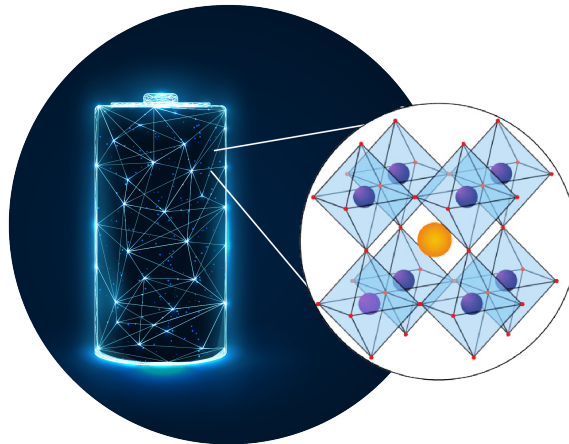


New Materials Development Using AI-ML Techniques

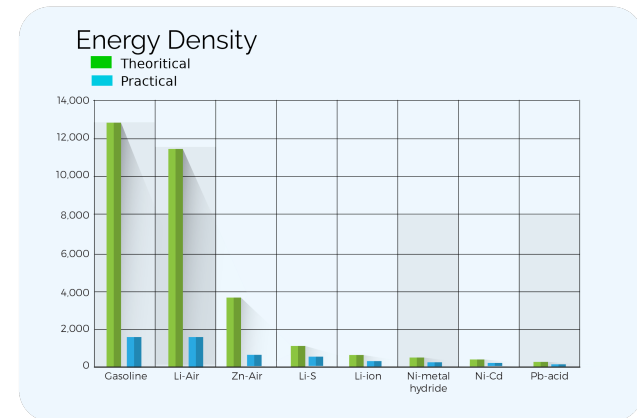
Sai Shanbhag,
Arjun Ravichandran
Data Science Team
Gyan Data Private Limited
Chennai-600113 India

Sukanya G,
Manjushri Rajaram
Project Associate
SENAI Lab, Dept. of Chemical
Engineering
IIT Madras



About the Client

The client operates as a subsidiary of one of the largest automobile OEM alliances. The client's organization has developed strong capabilities ranging from Advanced R&D, Advanced CAE (Computer-Aided Engineering), Product Development, Digital Vehicles and Information Systems Development. The client has a leading presence in electric cars and a reputed legacy in ICE vehicles.



Motivation

Given the high energy density profile of metal-air batteries, they offer exceptional practicality for auto and domestic applications. However, the battery's performance is super-sensitive to a pair of sluggish electrochemical reactions. As is true of most domains, materials selection is key in achieving the desired performance and scalability. Recent discoveries of novel electrocatalysts indicate enormous potential to improve the battery's energy efficiency.

Problem

The client's research and advanced engineering team was interested in developing metal-air batteries for novel applications. Since the number of possible electrolyte-electrocatalyst pairs can range up to millions, identifying the best performing pair is no easy feat. Gyan Data proposed a machine learning methodology to fast-track this materials discovery process.

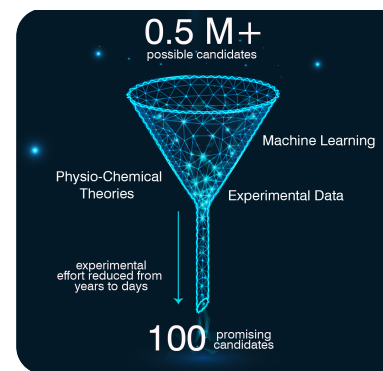
Descriptor generation process

Development of electrocatalysts is a research intensive process. Large open repositories like the Materials Project and Open Quantum Materials Database are the defacto standards for obtaining data. However, performance data of known perovskites is limited to experimental results reported in scientific literature. A tailored database of descriptors and catalytic activities was compiled after an extensive literature survey. Descriptor data for crystal structure, covalency, exchange interaction, electron occupancy and charge transport was also generated.

ML algorithms training

Data from hundreds of scientific reports was normalized to reference electrode - LaCoO_3 and LaMnO_3 electrocatalysts. The activities from multiple sources with

different measurement units were scaled for uniformity. The compositionally distinct and breakthrough electrocatalysts were intentionally held out for final model validations. Multiple machine learning algorithms were trained on the remaining data. Learning hyperparameters were tuned in a K-fold cross validation scheme. The resulting best model was interpreted in terms of the most influential descriptors. The predictive accuracy for breakthrough electrocatalysts was analysed. These results were verified with the client electrochemist experts and confirmed to be in agreement with their past experience.



Funnelling best-in-class electrocatalyst-electrolytes

Descriptor data was generated for half-a-million possible perovskites. The resulting activity predictions for both OER and ORR were persisted into a database for funnelling exercise by the client and for high-dimensional visualizations. Top 100 high activity perovskites were suggested to the client using a custom routine for synthesis and characterization. Having adopted the methodology, the client is intent on extending it to other materials systems for fast-tracking discovery.