

Research Review Presentation on

Building a Better Battery with Data-Driven Techniques

Submitted By

Prateek Gandhi
20SD06011

Under the Supervision of

Dr. B. Pattabhi Ramaiah



School of Mechanical Sciences
Indian Institute of Technology, Bhubaneswar

September 2021

Building a Better Battery with Data-Driven Techniques

Prateek Gandhi
20SD06011

Abstract:

Organic solvents are used as electrolytes in lithium-ion batteries currently. They work well by most accounts, lithium flows through them pretty quickly, they're manufactured at scale but organic solvents are a problem because they're very flammable. Hence there is a need for entering the unknown part of material science where we are combining what we know from the conventional approaches with these new ideas from big data and machine learning to accelerate the current work of search for materials for solid electrolytes in the lithium-ion batteries. The development of new solid electrolytes could ease a lot of concerns over the safety, stability, energy density, and cycle life of commercial batteries. Furthermore, the development of new solid electrolytes could facilitate the development of structural batteries for the weight and volume delicate applications of electric spacecraft and aircraft.

Keywords: Solid Electrolyte, Machine Learning, Lithium-ion batteries, Superionic Conductors, Solid-State Batteries.

CONTENTS

Contents	Page No.
INTRODUCTION	1
1.1 Theoretical background	1
1.2 Working of a battery	1
1.3 Scope of improvements	2
1.4 Electrolytes	3
REVIEW OF THE TOPIC	3
2.1 Literature review	3
2.2 Major focus areas	5
2.3 Model Formulation	7
2.4 Underlying Mathematics in Logistic Regression	9
2.5 Important findings	11
2.6 Comparison of performance of Scientists and Machine	13
CONCLUSION/ SUMMARY	15
3.1 Using data to guide tomorrow's efforts	15
3.2 Simultaneous optimization of conductivity and stability	16
3.3 Conclusions	19
3.4 Interpretation	21
3.5 Summary	22
REFERENCES	22

1. INTRODUCTION

1.1 Theoretical Background

Environmental and energy-based issues have become the major areas of concern in today's day and age. The search for alternative sources of energy is ongoing. The energy economy, at present, which is mainly based on fossil fuels, is at risk due to the decrease in non-renewable resources and the continuously increasing demand for energy. To add to it, CO₂ emissions associated with the use of fossil fuels are one of the main genesis of global warming, which is becoming an important issue at the global level.[1]

Investments for the utilization of renewable energy resources are growing worldwide, with particular attention to solar, wind and battery power systems. Batteries have many advantages as a substitute source of energy storage mechanisms. At present, the conventional battery technologies, such as lead-acid and nickel-cadmium, are steadily being replaced by lithium-ion batteries, fuel-cell technologies and nickel-metal hydride batteries. Li-ion battery stands as a precursor and market leader when compared to other energy systems.

Lithium-ion batteries, in particular, have enabled major innovations across the spectrum of energy technologies. In the automotive sector with Tesla. Other companies are thinking about heavy equipment being powered by batteries like aeroplanes and of course, mobile electronics. So it is safe to say that Lithium-ion batteries are run the gamut of powering all kinds of technologies.[1]

1.2 Working of a battery

The cross-section of a cylindrical battery can be visualized as having a structure similar to a roll of paper that is connected to the positive and negative terminals. The cell is made up of several different layers. First, we have a copper foil which acts as the current collector for the anode then we have a porous anode which is normally made up of graphite, on top of that we have a microporous separator which prevents the anode and cathode from touching each other. Then we have a porous cathode which is made up of a transition metal oxide and an aluminium foil which acts as the cathode current collector in each of the porous phases. We also have a liquid electrolyte that allows for the movement of lithium ions. When the battery is fully charged, lithium exists in between the carbon layers of the graphite. When the battery is discharged the lithium becomes a lithium-ion and an electron. The lithium ions can travel

through the electrolyte towards the cathode however the electrons have to go around an external circuit where useful electrical work can be extracted. At the cathode, the lithium ions and electrons recombine and go into the layers of the transition metal oxide in a process called intercalation. When charging the battery the opposite reaction occurs moving the lithium back into the anode. This back and forth motion coins the term “rocking chair mechanism” which represents how a lithium-ion battery works.[17]

1.3 Scope of improvement

These batteries aren't perfect. There are four main areas where battery technology can get a lot better. Safety, of course, should be the top priority. The safety of batteries needs to be unquestioned and it's very questioned presently. Other factors like energy density always need to come up, cost always needs to come down and cycle life needs to improve. For electric vehicles where the user would want to be able to drive for 10 or 20 years without replacing the battery, it should have an excellent cycle life. The ways that these metrics can be improved are being explored currently. Whilst there's been year-on-year improvements to batteries, there still is a need for significant improvements to achieve mainstream adoption. One of the key metrics is cost, where the frequently cited number to be cost comparable with the internal combustion engine is 100\$/kW-h at the pack level which means we need a cell cost to be about 50\$/kW-h. Currently, this is just over 200\$/kW-h at the pack level and just over 100\$/kW-h at the cell level.[17] Though this varies depending on the volume and application. Next, we need to improve the energy density, which represents the range of an electric vehicle, from approximately 250 W-h/kg to about 500 W-h/kg. Also, there is a need to increase the power density which represents the acceleration of the electric vehicle from current levels of about 3 kW/kg to 12 kW/kg.[17] However, as we do this we shouldn't forget about safety as we add more and more energy, it becomes more dangerous and thus there's a need to eliminate the threat of thermal runaway. Lifetime is also critical especially as we aim to apply these batteries in the large-scale grid, energy storage applications and the aim is to hopefully increase the lifetime from about 8 years to 15 years. If we want to electrify more applications, the temperature range over which cells operate must also be improved and as we manufacture more and more of these cells we should also aim to improve their predictability and consistency. Finally, as we start to produce many millions of these cells the sustainability of the system also needs to be considered with the recyclability generally quite low at the

moment but with a need to move to near 100% recyclability of the materials in the near future.

1.4 Electrolytes

One of the aspects of battery technology that is at the heart of a lot of these problems are the electrolytes used. The battery has an anode and a cathode, they are the two terminals and then there is an electrolyte in the middle that separates the two terminals and it mediates the transport of lithium back and forth when the battery is charged and discharged. In the cells of the smartphone, of the laptop, the electrolyte used is an organic solvent, which is liquid. Aqueous or water-based solvents can't be used because lithium reacts with water. So organic solvents are used. They work well by most accounts, lithium flows through them pretty quickly, they're manufactured at scale but organic solvents are a problem because they're very flammable. Example: nail polish remover. So if the battery is overheated or punctured, it could catch fire and that is going to be a huge issue. But aside from that, they're also not great because they are very reactive so a large voltage across a liquid electrolyte can't be put because it'll start to vaporize and break down. They are also low density and they don't suppress dendrites and dendrite growth is one of the main causes of early battery death.[5]

2. REVIEW OF THE TOPIC

2.1 Literature Review

The research is headed towards replacing the liquid electrolyte with the solid material that can do the same job and that hopefully has all the benefits of the liquids without these major drawbacks. Solids are inherently more stable and less reactive. They have higher density so smaller devices could be made potentially and they're very rigid so they can suppress this growth of dendrites and potentially lead to increased cycle life. So solids can help on all the metrics that were originally discussed. Therefore, solid electrolytes are fascinating and a very promising way to go to improve battery technology.[5]

It's been over 40 years and research is still going on on this problem because solid electrolyte materials that work in batteries have to satisfy a bunch of criteria at the same time as shown in Figure 1.

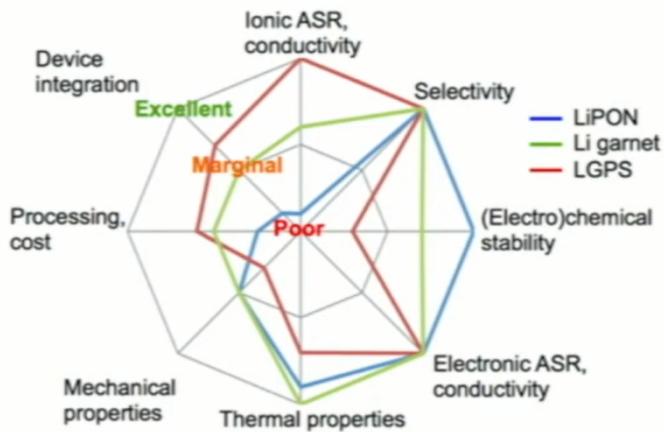


Fig. 1: Conditions on which potential candidate materials are being searched.^[3]

There is conductivity and stability as well as processing ease from the manufacturing scale point of view. So there are a lot of requirements that should be met at once and if searching for promising materials is done based on trial and error, then the odds of finding a material that is going to satisfy ten criteria at once is going to be extremely low since predicting ionic conductivity of any material is very hard.[5]

From the 70s to the present we have generated a lot of data of XY pairs where X is a material that was picked by a random guess and Y is the property which was being measured, ionic conductivity for example. So now the idea is to reformulate this as let's not just keep trying things but let's take this data that's already out there and let's reformulate this as a problem of supervised learning where we are taking the X and the Y's we know and look for the function f which can map between the two of them and in the material science world this is a structure-property relationship.[5]

The question which is driving the research is can we learn effectively from the data that we have already generated without having to generate a whole bunch more. And machine learning is the way to do that to hopefully go beyond trial and error.

There are all kinds of different combinations of elements, which are close to infinite and hence this becomes just totally intractable. That is where the power of machine learning comes in so in the landscape of the accuracy of a scientific approach versus the size of the

searchable space, machine learning is going to be less accurate than DFT (density functional theory simulations) but orders of magnitude faster as shown in figure 2.

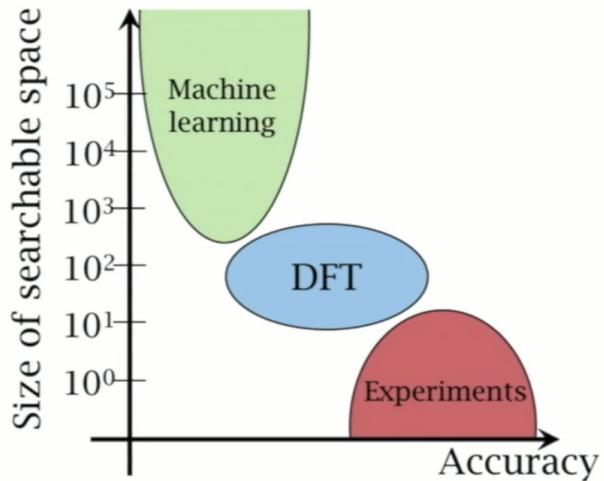


Fig 2: Comparison of ML, DFT and Experiments on size of searchable space and accuracy. [4]

So it is easy to get through billions of candidate materials. This will accelerate the development of solid-state batteries as it beats trial and error.

2.2 Major Focus Areas

With the model and results presented herein, a new approach for the discovery of solid electrolytes with the potential to identify the next generation of high-performance Lithium-ion battery solid electrolytes from thousands of candidates is presented. The development of new solid electrolytes could ease a lot of concerns over the safety, stability, energy density, and cycle life of commercial batteries. Furthermore, the development of new solid electrolytes could facilitate the development of structural batteries for the weight and volume delicate applications of electric aircraft and spacecraft. Additionally, this screening approach allows for the inclusion of additional requirements, including environmental concerns like earth abundance.[5] The impact of a highly stable new solid material with liquid-like lithium conduction cannot be overemphasized: solid-state Lithium-ion batteries stand to improve on the safety, performance, and lifetime of the energy storage technology, and in doing so help to realize an electrified future with less dependence on fossil fuels.

2.3 Experimentation / Model Formulation

Experimentation:

Experimental measurements of ionic conductivity at room temperature for 40 different materials are gathered and these are all crystalline materials and they have well-defined structures that have been characterized and so for all 40 of them, the researchers came up with 20 features which are essentially pieces of information that they think might have some bearing on predicting ionic conductivity. Some of them came from the literature, from things that people have proposed, some of them came from intuition.[5]

The first thing that they did was that they just computed all twenty of these for the forty materials and they looked at them one by one just to see which one predicts conductivity and the correlation coefficients are shown in figure 3 and we can see that they are all small. They are all less than plus or minus 0.3.

Composition	RT bulk ionic conductivity (S cm^{-1})	Feature	Pearson correlation coefficient
$\text{LiLa}(\text{TiO}_3)_2$	1×10^{-3}	1 Volume per atom ^a	0.20
$\text{Li}_{0.81}\text{Sn}_{0.81}\text{P}_{2.19}\text{S}_{12}$	5.5×10^{-3}	2 Standard deviation in Li neighbour count	0.22
$\text{Li}_{10}\text{Ge}(\text{PS}_3)_2$	1.4×10^{-2}	3 Standard deviation in Li bond ionicity	-0.04
$\text{Li}_{10.35}\text{Si}_{1.35}\text{P}_{1.65}\text{S}_{12}$	6.5×10^{-3}	4 Li bond ionicity ^a	-0.18
$\text{Li}_{14}\text{ZnGe}_4\text{O}_{16}(2)$	1.0×10^{-6}	5 Li neighbour count ^a	-0.19
$\text{Li}_3\text{Ca}(\text{NH})_2$	6.4×10^{-6}	6 <i>Li-Li bonds per Li</i> ^a	0.06
$\text{Li}_3\text{Ge}_3\text{O}_{15}$	5.0×10^{-6}	7 <i>Bond ionicity of sublattice</i> ^a	-0.28
Li_2NH	2.5×10^{-4}	8 Sublattice neighbour count ^a	-0.13
Li_2S	1.0×10^{-13}	9 Anion framework coordination ^a	-0.06
$\text{Li}_{13.6}\text{Si}_{2.8}\text{S}_{1.2}\text{O}_{16}$	6.0×10^{-7}	10 Minimum anion-anion separation distance ^a (\AA)	0.09
$\text{Li}_{14}\text{Ge}_2\text{V}_2\text{O}_{16}$	7.0×10^{-5}	11 Volume per anion (\AA^3)	-0.01
$\text{Li}_{15}\text{Ge}_3\text{V}_2\text{O}_4$	6.03×10^{-6}	12 Minimum Li-anion separation distance ^a (\AA)	0.20
$\text{Li}_{14.8}\text{Ge}_{3.4}\text{W}_{0.6}\text{O}_4$	4.0×10^{-5}	13 Minimum Li-Li separation distance ^a (\AA)	-0.10
$\text{Li}_3\text{Fe}_2\text{P}_3\text{O}_{12}$	1.0×10^{-7}	14 Electronegativity of sublattice ^a	-0.16
Li_3N	5.75×10^{-4}	15 Packing fraction of full crystal	0.16
Li_3P	1.0×10^{-3}	16 Packing fraction of sublattice	0.19
$\gamma\text{-Li}_3\text{PS}_4$	3.0×10^{-7}	17 Straight-line path width ^a (\AA)	0.07
$\text{Li}_3\text{Sc}_2\text{P}_3\text{O}_{12}$	1.0×10^{-10}	18 Straight-line path electronegativity ^a	-0.29
$\beta_{11}\text{-Li}_3\text{VO}_4$	4.4×10^{-8}	19 Ratio of features (4) and (7)	-0.03
$\text{Li}_3\text{B}_2\text{O}_{12}\text{Cl}$	1.0×10^{-7}	20 Ratio of features (5) and (8)	-0.18
$\text{Li}_4\text{BN}_3\text{H}_{10}$	2.0×10^{-4}	Constant term	—
$\gamma\text{-Li}_3\text{GeO}_4$	3.1×10^{-12}		
Li_3SiO_4	2.4×10^{-10}		
$\text{Li}_3\text{La}_3\text{Bi}_2\text{O}_{12}$	2.0×10^{-5}		
$\text{Li}_3\text{La}_3\text{Nb}_2\text{O}_{12}$	8.0×10^{-6}		
$\text{Li}_3\text{La}_3\text{Ta}_2\text{O}_{12}$	1.5×10^{-6}		
Li_3Ni_2	1.5×10^{-7}		
$\text{Li}_3\text{BaLa}_2\text{Ta}_2\text{O}_{12}$	4.0×10^{-5}		
Li_6FeCl_8	1.0×10^{-4}		
Li_6NBr_3	1.5×10^{-7}		
$\text{Li}_3\text{SrLa}_2\text{Ta}_2\text{O}_{12}$	7.0×10^{-6}		
$\text{Li}_7\text{La}_3\text{Zr}_2\text{O}_{12}$	3.5×10^{-4}		
$\text{Li}_3\text{P}_2\text{S}_{11}$	4.1×10^{-3}		
LiAlH_4	2.0×10^{-9}		
LiAlSiO_4	1.4×10^{-5}		
LiBH_4	2.0×10^{-8}		
LiI	1.0×10^{-6}		
LiNH_3	4.0×10^{-10}		
$\alpha'\text{-LiZr}_2\text{P}_3\text{O}_{12}$	5.0×10^{-8}		

Fig 3: 40 experimental measurements of conductivity and 20 features for predicting ionic conductivity. [5]

So out of their initial twenty guesses, none of them really were good predictors. So they have used an axiom in machine learning which is that if we have a bunch of weak learners or weak

descriptors, we can put them together in a creative way that we can create a strong learner or a better predictor.[5] So even though there's no single feature with a strong correlation across this diverse set, we hope that we can creatively put these together in a way that gives us that predictive power.

Model Formulation:

To move forward we have to choose a model that we want to use. A binary classifier model was chosen which just predicts essentially a yes or a no. So the training set was broken down into good conductors and bad conductors with a cut-off of 10^4 . So that is the lowest we can have and still have it be considered a good electrolyte material. This splits this up into about 1/4 good conductors about 3/4 bad conductors. There are no conventions in this sort of space especially in machine learning with this small of a data set but after testing other models, this was the one that the researchers found worked the best as shown in Eqn. 1.[5]

$$P_{superionic}(x) = \frac{1}{1 + \exp(-\theta^T x)} \quad (1)$$

In the logistic regression, it has θ transpose x in the denominator. x 's are the combination of features and the θ 's are the regression coefficients. So the question is now which of those 20 features would give us the best predictive model for that data that we're training it on. We take the 20 candidate hypotheses and the 40 data points and then we want to find the best feature set we want to fit the data to. Having a small data set usually is bad because we can have problems with overfitting and it is harder to deal with. But one of the benefits is that with a small data set we can search the entire model space in a reasonable amount of time. We then measured the cross-validation error of the model which tells about how the model is going to predict unseen data and that's what the solid red line represents in figure 4. This is plotted against the number of features. From the figure, there is a minimum in the cross-validation error at five features.[5]

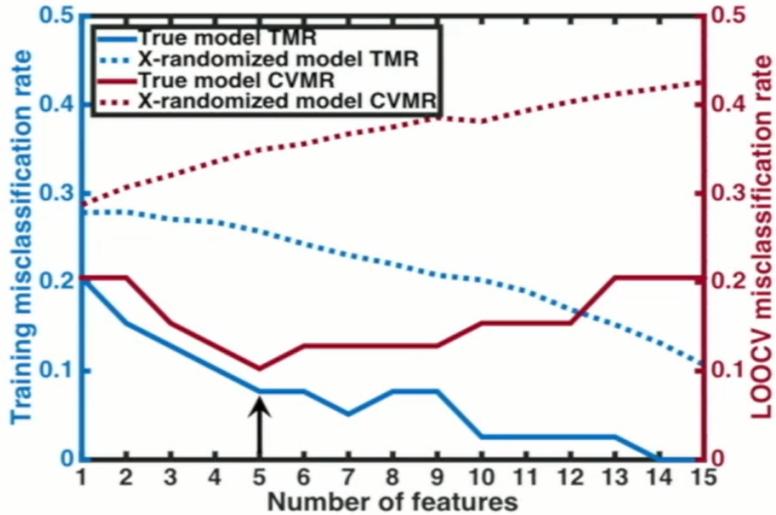


Fig 4: Cross-validation to determine optimal features. [5]

That means that there is one model out of this one and a half million which gives us the least error and it's the best model that we could find with this data set. The dotted red line is where we replaced all that feature data with 20 columns of random noise that was just generated by a random number generator and the fitting process was redone. The error is about three times lower when we're using the real features than when we're using noise and hence it serves as a confidence builder that we are building this model on real information and we're capturing the result which is statistically significant here. The hope is that this optimal model would give an accurate guess about 90% of the time. The model developed is shown in Eqn. 2.

$$\theta^T x = 0.817LLB - 1.323SBI - 1.028AFC + 2.509LASD - 1.619LLSD - 1.944 \quad (2)$$

Where *LLB* is Lithium-lithium bond number, *SBI* is Sublattice bond ionicity, *AFC* is Anion framework coordination, *LASD* is Li-anion separation distance and *LLSD* is Li-Li separation distance. These five features are easily interpreted in a material science context. This equation is based on five pieces of information and this takes, if we set it up an automated script, less than a second like a microsecond to compute, whereas it would have taken one week to get an estimate of conductivity with density functional theory simulations which use physics and have to solve the Schrodinger equation and simulate atoms. Hence, this is a huge speed improvement.[5]

2.4 Underlying Mathematics in Logistic Regression:

When the dependent variable is binary(dichotomous), Logistic regression is the appropriate regression analysis to conduct. A binary variable just means a variable that has only 2 outputs, for example, a person will survive this accident or not, the woman is pregnant or not. The outcome can either be yes or no (2 outputs). This regression technique is similar to linear regression and can be used to predict the Probabilities for classification problems.

Logistic regression is predictive analysis, just like all regression analyses. It is used to describe data and to explain the relationship between one dependent binary variable and one or more ordinal, nominal, interval or ratio-level independent variables.

Logistic Regression is another statistical analysis method borrowed by Machine Learning.

Now, how does logistic regression squeezes the output of linear regression between 0 and 1? Well, there's a little bit of math included behind this and it is pretty interesting. The logistic function is given by:

$$P = \frac{1}{1+e^{-(\theta_0 + \theta_1 x)}}$$

How similar it is to linear regression?

We all know the equation of the best fit line in linear regression is:

$$y = \theta_0 + \theta_1 x$$

Let's say instead of y we are taking probabilities (P). But there is an issue here, the value of (P) will exceed 1 or go below 0 and we know that range of Probability is (0-1). To overcome this issue we take "odds" of P:

$$\frac{P}{1-P} = \theta_0 + \theta_1 x$$

Furthermore, we know that odds can always be positive which means the range will always be $(0, +\infty)$. Odds are nothing but the ratio of the probability of success and probability of failure. Now the question comes out of so many other options to transform this why did we only take 'odds'? Because odds are probably the easiest way to do this, that's it.[18]

The problem here is that the range is restricted and we don't want a restricted range because if we do so then our correlation will decrease. By restricting the range we are decreasing the number of data points and of course, if we decrease our data points, our correlation will

decrease.[18] It is difficult to model a variable that has a restricted range. To control this we take the log of odds which has a range from $(-\infty, +\infty)$.

$$\log\left(\frac{P}{1-P}\right) = \theta_0 + \theta_1 x$$

This was 80% of the maths. Now we just want a function of P because we want to predict probability. To do so we will multiply by exponent on both sides and then solve for P.

$$\exp[\log(\frac{P}{1-P})] = \exp(\theta_0 + \theta_1 x)$$

$$e^{\ln[\frac{P}{1-P}]} = e^{\theta_0 + \theta_1 x}$$

$$\frac{P}{1-P} = e^{\theta_0 + \theta_1 x}$$

$$p = e^{(\theta_0 + \theta_1 x)} - p e^{(\theta_0 + \theta_1 x)}$$

$$p = p\left[\frac{e^{(\theta_0 + \theta_1 x)}}{p} - e^{(\theta_0 + \theta_1 x)}\right]$$

$$1 = \frac{e^{(\theta_0 + \theta_1 x)}}{p} - e^{(\theta_0 + \theta_1 x)}$$

$$p[1 + e^{(\theta_0 + \theta_1 x)}] = e^{(\theta_0 + \theta_1 x)}$$

$$p = \frac{e^{(\theta_0 + \theta_1 x)}}{[1 + e^{(\theta_0 + \theta_1 x)}]}$$

$$p = \frac{1}{[1 + e^{-(\theta_0 + \theta_1 x)}]}$$

Now we have our logistic function, also called a sigmoid function. The graph of a sigmoid function is as shown below. It squeezes a straight line into an S-curve.[18]

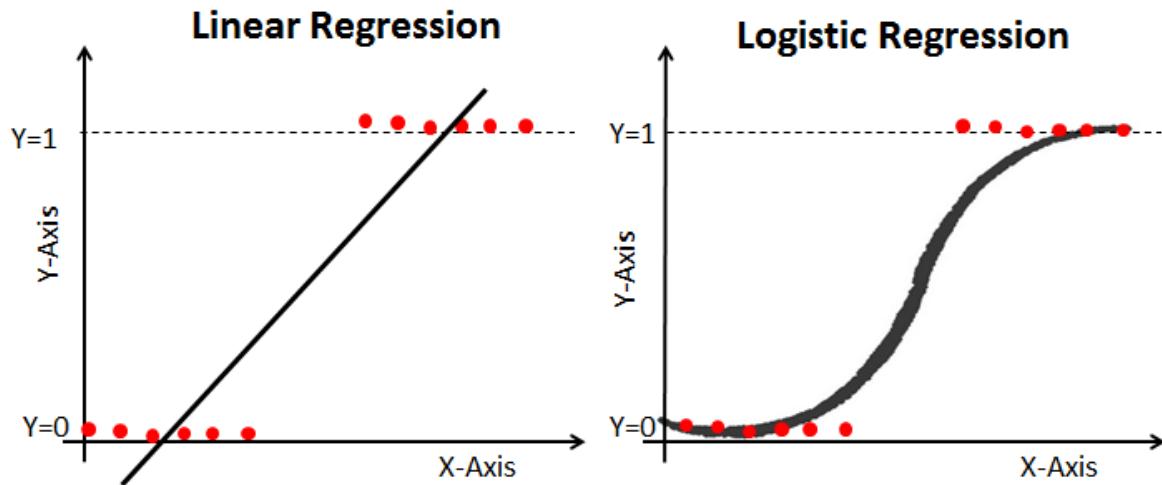


Fig 5: Linear Regression vs Logistic Regression [18]

2.5 Important Findings

Over 12,000 candidate materials were found in the online database of the materials project which is curated at Lawrence Livermore National Laboratory. They also included six extra screening metrics which weren't built from machine learning. These were built on simple heuristics. So when we take these twelve thousand candidate materials we first apply these six screening metrics, and this gets the number of candidate materials to reduce down to around 300 materials. 300 materials are still too many for researchers to do 300 experiments or 300 simulations so this is where we apply the conductivity predictor. After applying it to those 300 materials, we're able to get it down to 21 and hence we're able to do the first multi-property screening of all known lithium-containing compounds for solid electrolyte materials because we have this conductivity model which is so much faster than what we had before.[5]

Out of those 21, we found about half of them showed really exciting levels of lithium-ion conduction at room temperature and we found this through doing these physics-based simulations.[5] Figure 6 is a mean squared displacement plot for how far lithium migrates in time in these different materials.

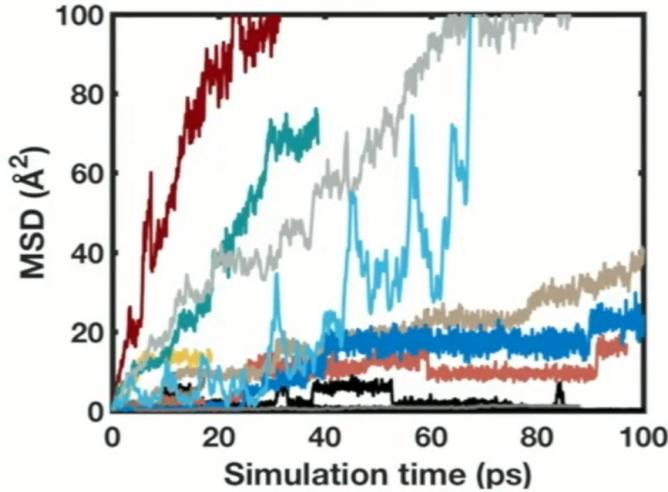


Fig 6: Mean squared displacement plot of Lithium migration for 10 materials. [5]

This is a common way of showing and tracking how far lithium migrates in time in these different materials. Since these physics-based codes are really slow we have to operate on a picosecond time scale. The slopes of these curves are what tell us about conductivity which is our original metric. We can break these materials down into three categories so we have four materials here that are on par with the best-known materials with a conductivity of about 10^{-2} Siemens per centimetre which is exciting. We then have materials with a conductivity of 10^{-4} , which is near the cutoff of good materials. Then we get several materials that don't show any motion at all and so this gives a 50% hit rate for the model success rate.

If we compare it to the incumbent methods, this model's success rate is between 35 to 50 per cent.[5] The same procedure was repeated but this time the simulating materials were chosen at random and what we found, in that case, was that the success rate was about 10%. So we can see that we have about a three to five times improvement in the rate of success. The histogram in figure 7 shows that we have three categories - bad conductors, medium conductors and good conductors.

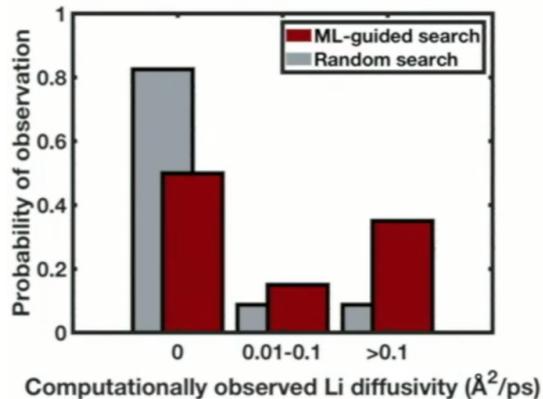


Fig 7: Comparison of ML guided search and Random search. [6]

The grey bars are what we find with random searches and then the red bars are what we find with machine learning guided searches. We can see that we get about four times more, good conductors, with ML guided searches. Therefore there is a tangible improvement that's offered here by machine learning. This 3x improvement came from a model that was trained on just 40 data points.[6] Therefore, it's worth pointing out that even small amounts of data can give real tangible improvements in the way that we discover materials.

2.6 Comparison of performance of Scientists and Machine:

Researchers conducting this study took a list of 25 materials and gave it to a group of PhD students in the material science department of Stanford that are trained in electrochemistry essentially and then they gave that same list of materials to the code to the machine learning algorithm and they said: “find us the good lithium-ion conductors in this batch of materials”.[7]

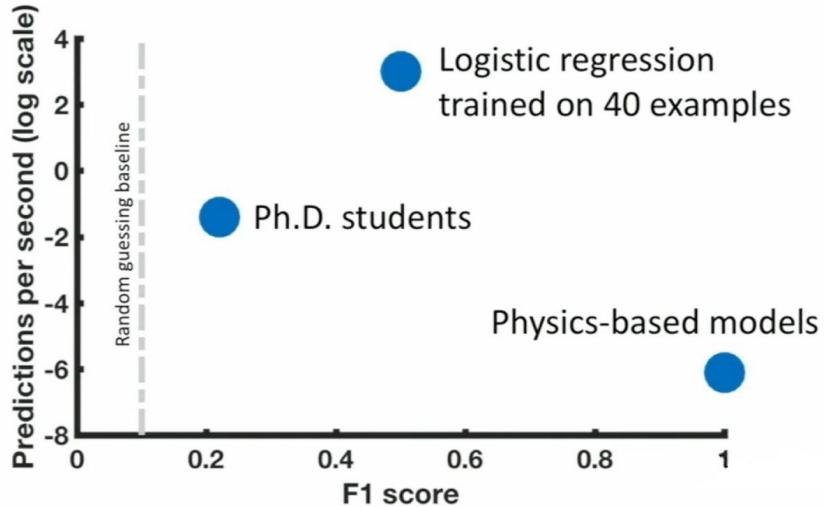


Fig 8: Performance of algorithm compared to traditional methods [7]

So we have two axes here in figure 8, F1 Score which is basically the accuracy of the predictions and we have speed on the left. The dotted line is the random guessing baseline so that there's about 10% of those materials are our good conductors that's what we would expect to see if you were guessing at random and the PhD students perform better by double than random guessing - basically, they have an accuracy that's twice as high which is great. So they're definitely beating random guessing by a factor of two.[7] So, scientists are not truly guessing at random. But then the logistic regression model trained on 40 examples beats random guessing by a factor of about five and beats the students by about two to three times and if that's not good enough we have this is a log scale on the left. It takes about a million times or it's a million times faster for the machine learning model to make those predictions so students would take about a minute and the computer takes less than a second. For reference, we have these physics-based models which are used to validate the predictions we assume these are 100% accurate so that's why they're up here way on the right but they are incredibly slow so each material takes several weeks to get a good validation.[7] So this is a sort of our Bermuda Triangle of machine learning and students guessing materials. So what we've learned from this is machine learning offers several times or guessing that several times more accurate than trial and error.

3. CONCLUSION/ SUMMARY

3.1 Using data to guide tomorrow's efforts:

So two questions that are of importance after doing this work are: one of them is where do we look to find the promising outliers, the good iron conductors and can we quantify the likelihood of finding them?

That's difficult to answer most times like what is the likelihood that I'm going to make a major discovery. That's a difficult thing to answer especially with first-principles methods and it'd be great if we could say something about you know how likely am I to make a discovery based on if I'm looking here versus if I'm looking there. Electrochemical stability is also a very important metric in these materials. We talked about conductivity so far, so conductivity is kind of indicative of power density - it has to do with how much current you can get out of the cell. The electrochemical stability window tells you about the voltage you can have across the cell so that's more like energy density. So optimizing both of these at the same time is a really difficult challenge. Getting good iron conductors is hard enough but getting good ion conductors that are electrochemical state below for a wide range of voltages is even harder.[4-7, 19]

For an electrolyte, in a conventional battery, we put three demands on it if we move beyond conductivity. We also need it to be stable against oxidation at the cathode because at the interface it tends to be oxidized. Also, we need it to be stable against reduction as well as it is touching anode and hence it will tend to be reduced. So we are looking for a material that can satisfy all of these demands at once. It's a very tough optimization to do.[6]

One method that has been proposed to make this optimization a little bit easier is to split the electrolyte up into two electrolytes as shown in figure 9.

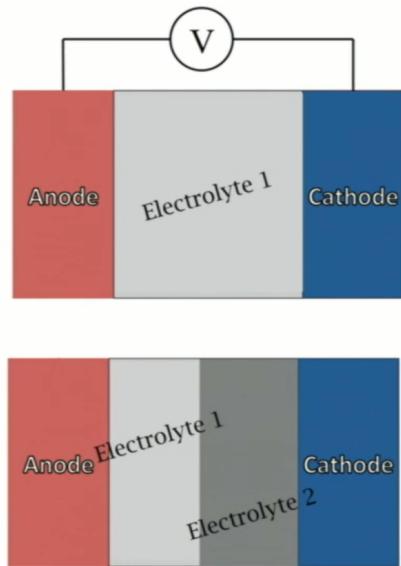


Fig 9: Splitting up the electrolytes to ease the optimization [6]

Now both of them need to have good ionic conductivity but electrolyte one is touching the anode and not the cathode so it only needs to be stable against reduction. Electrolyte two is only touching the cathode and not the anode so it only needs to be stable against oxidation. Therefore by using this method the demands can be reduced on these materials but then this method introduces additional difficulties. Manufacturing this configuration is hard. Getting the materials to stack right on top of each other with minimal defects and less interfacial resistance is tough. It is not clear which strategy is better since this has been something that's more recently thought about.[6]

3.2 Simultaneous optimization of conductivity and stability:

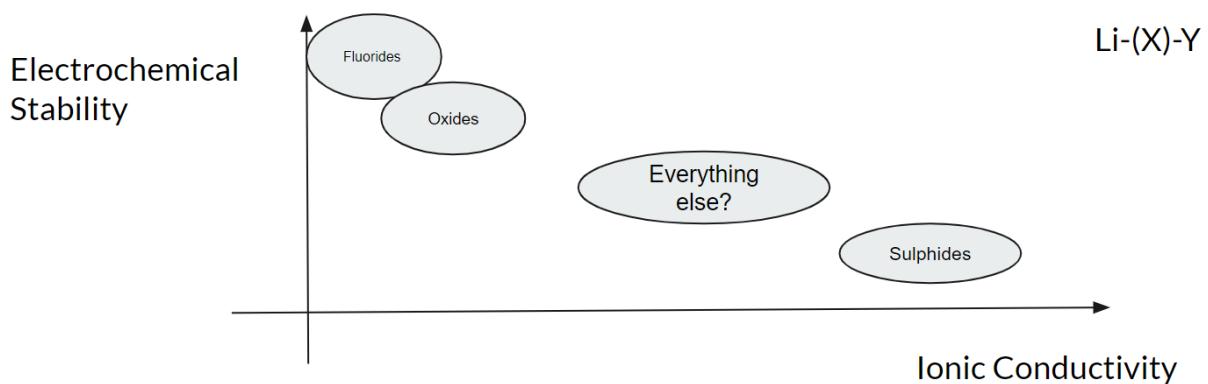


Fig 10: Electrochemical Stability vs Ionic Conductivity of Anions [19]

Figure 10 is our map and we want to know that where do the materials lie in this space and one trend that is sort of known out there. Now, we can write any material like Li-(X)-Y. It has lithium in it, has a bunch of other junk and then there's an anion in there somewhere like oxygen or fluorine and it's known that that anion is pretty predictive of where these materials will lie in this space.[19] Fluorides and oxides tend to be poorly conductive but pretty stable whereas sulfides tend to be good conductors but also not very stable and these trends have sort of emerged in recent years just through a few different data points for each family. Everything else - is it hasn't been looked at as extensively as these three families.

We now want to quantify this trend and provide a little bit of insight. This is very difficult to do with first principles so if all that you have in your toolkit is the Schrodinger equation, it would be a very difficult problem but because we have these models which allow us to generate a lot of predictions very quickly we can now answer this question with data.[19] So we take our landscape here we can just compute conductivity now or predict it and we can use some DFT stability methods to compute stability for 12-thousand materials and we can throw it up on this plot and kind of look at where the data lies and see if it's telling us anything.

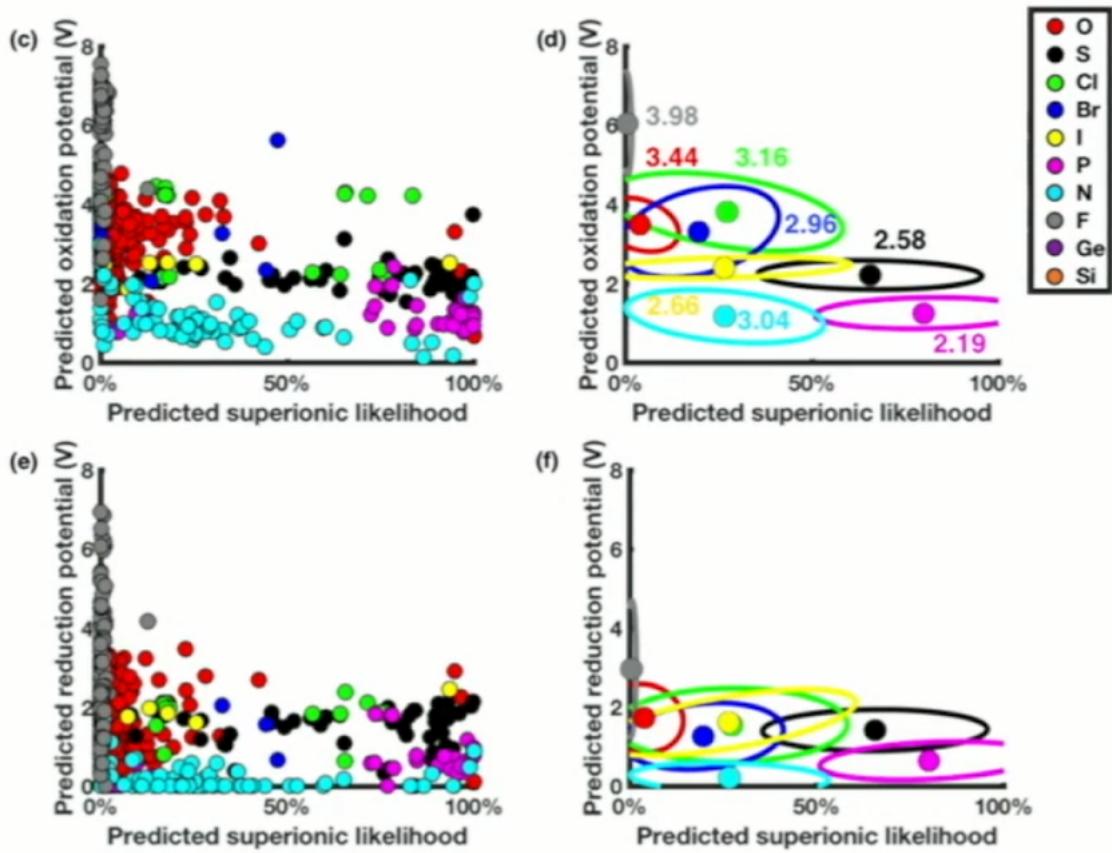


Fig 11: Probability distribution of different anions [19]

So we took our conductivity model and we took existing electrochemical stability analysis methods and we computed the conductivity on the x-axis with oxidation potential and reduction potential and we plotted it for over a thousand materials and broke it up by that anion in the structure. So all these red dots as shown in figure 11, are the oxides. The grey ones are fluorides, and the black ones here are sulfides.[19] So we can see it kind of follows that trend that the community has deduced from the limited data that we have. This trend is also more or less true with reduction as well as with oxidation. To clear things up a little bit we just replace the data by essentially the covariance. These ellipses designate one standard deviation of the data in all directions and then we can say that the data that's been generated has come from underlying probability distributions that look a lot like this and if we're going to keep sampling materials and reporting them in the literature it's probably going to come from that same distribution. Essentially what we've done in the past is probably indicative of what we're going to do in the future and we can quantify the likelihood that we get up there

by using statistical methods such as integrating Gaussian distributions and finding cumulative distribution functions of this regime of outlier space.[7]

$$P(\text{promise}|\text{species}) = \frac{\int_{0.5}^1 dP_{LR} \int_1^{10} dE_{gap} \int_4^{10} dV_{ox,max} e^{-\frac{1}{2}(\vec{x}-\vec{\mu})^T \Sigma^{-1} (\vec{x}-\vec{\mu})}}{\int_0^1 dP_{LR} \int_0^{10} dE_{gap} \int_0^{10} dV_{ox,max} e^{-\frac{1}{2}(\vec{x}-\vec{\mu})^T \Sigma^{-1} (\vec{x}-\vec{\mu})}}$$

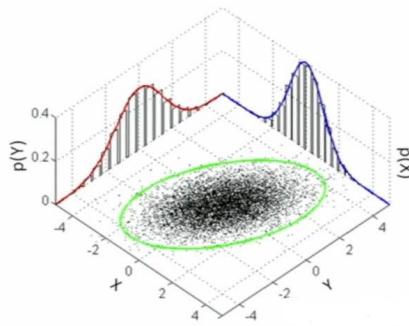


Fig 12: Using the past distributions to predict the future [7]

So we are fitting Gaussians to each of these families and we're saying the data for each family likely came from this probability distribution and if it did, how much of the tail of that distribution is in my promising outlier region because there are no data points there.[7] So we can't really count them but we can say what does the tail look like and we're essentially using the past to try to predict the future.

3.3 Conclusions:

The conclusions here that we found are that if we want a fast conducting stable material we were actually able to quantify that the oxides are, on an absolute scale, are pretty unlikely to yield these promising new materials and that's an interesting insight. Because a lot of research on this so far has been put on the development of electrolytes with oxide anion because they're very easy to make. But we found that oxides really on the scale of these materials are pretty low. The sulfides turn out, are also unlikely but are more likely than the oxides, about ten times more likely. But that's something else that we showed in that graph is really where things get interesting. So the things that are in between these sulfides and the oxides start to look really promising and the things that are in between are the halides, which are the chlorides, the bromides, the iodides - these are materials where there's not a ton of data present but the data that is there is pointing us towards something really interesting.[4-9, 19]

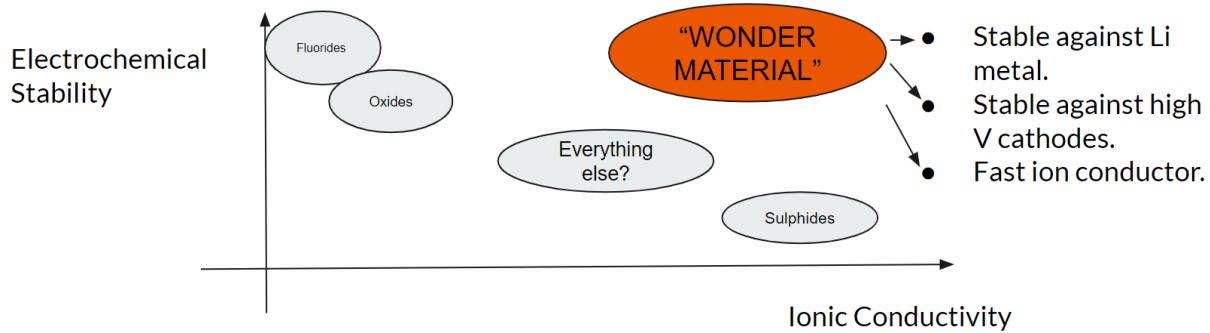


Fig 13: Likelihood of finding the Wonder Material [7]

What we know also is that certain families are better against reduction certain families are better against oxidation and so this has some implications on this final question we want to answer which is - how likely are we to find the “Wonder material” which satisfies all my criteria at the same time that is which gives us a lithium metal anode and which gives us a high-voltage cathode and gives us fast lithium-ion conduction. So coming back to the 1 versus 2 electrolyte system - if we want to build the lithium metal high voltage wonder material battery, we can do it with one electrolyte or we can do it with two. If we do with one, the probability of success here is the probability that, at least from the materials discovery perspective, we can find one material that satisfies all the requirements. But if we do it with two, the probability of success is the probability that we find one good conductor that is stable against oxidation times the probability that we find a good ion conductor that's stable against reduction and we now have quantitative numbers to do this calculation. So, for the two electrolyte case, we find that it's the likelihood of doing the first criteria is about 90% since there's a lot of good materials for that and the likelihood of the second one is 10% so the joint probability here is 9%. So that seems pretty good and the reduction stable material is a nitride so this 9% corresponds to fluoride and a nitride-based dual electrolyte.[4-9]

Now, if we want 1 electrolyte we don't get to mix a fluoride with the nitride, we have to choose one material. We have to find one material from one family that is likely to meet all these criteria and the chlorides are the most likely material family to do that. But the probability is just about 0.06 per cent and that's the best and the most likely of all of the options that we have. So, we can see that in going from a 1 to 2 electrolyte system we are 150 times more likely to discover good materials for this purpose. There is certainly more to this puzzle than just the materials. We have to think about processing and other factors, but the

first step is you have to find good materials for it and this is a quantitative roadmap that says we are 150 times more likely to find success if we start searching for this architecture than if you do by the single electrolyte system.[19]

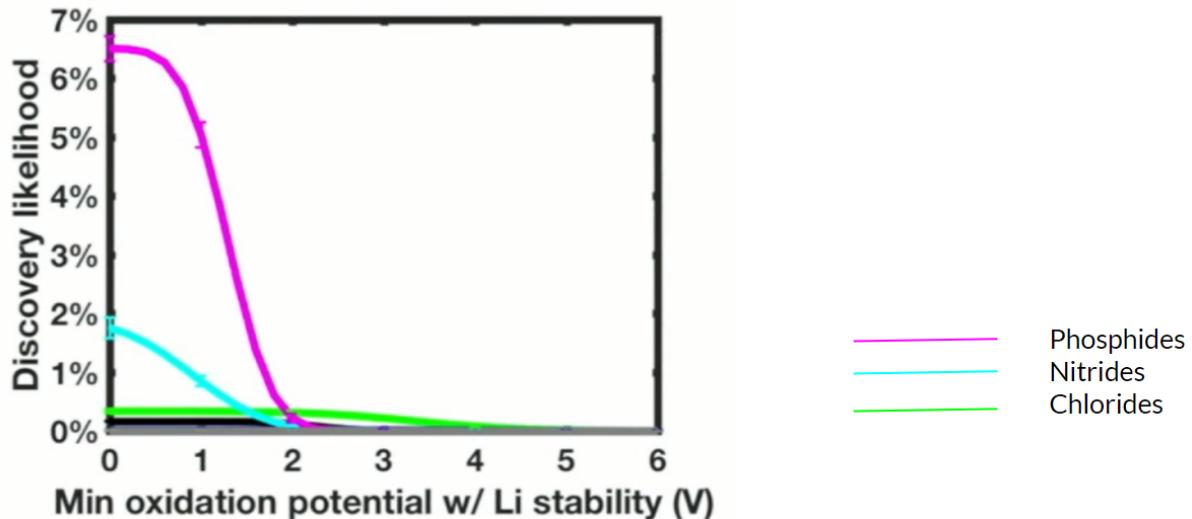


Fig 14: Discovery likelihood for different families for single electrolyte architecture [19]

Just to demonstrate that a little bit further, figure 14 shows the likelihoods for the various families - the likelihood that we can discover a promising outlier that is stable against lithium metal, a good ion conductor and is stable above a cathode potential of X. So pretty much everything goes to zero at about two volts so it is very unlikely that we find a material that will allow a more than 2-volt lithium metal battery and that defeats the whole purpose. The chlorides are the only ones that extend into this region of 4.5 which is where we need to be although the odds as we can see are essentially zero, so it's very small.[19]

3.4 Interpretation:

So, these insights are generated from models that are trained on data - that is an incomplete data set and we could say everything comes from this model that maybe works half the time so why should we believe this? It comes down to the way that we think about this as being the best we can do with the data that we have and the urgency of the energy challenge that we're faced with. We don't have a whole lot of time to continue in doing trial and error and it becomes really important for us to squeeze all predictive power out of the data that we can get and in the worst-case scenario that the models trained on this data don't have much predictive accuracy all that happens is that we're reduced to going back to trial and error

which is essentially what we were doing anyway. So we see this as the most informed way forward - could be wrong, but will only get better with time.[4-9]

3.5 Summary:

So we deployed machine learning to enable the first holistic screening of all known lithium-containing compounds which was about a million times faster than state-of-the-art methods, three times better than random guessing, and that gives us several new materials for development and also can give us some of these high-level insights on different strategies and how likely they are to succeed that are difficult to get from physics-based approaches. The really exciting thing is thinking about how important forty data points can be and if we go forward in the future and commit ourselves as a research community to cataloguing this data religiously, it's really exciting to think about what 500 data points would be able to get us or what a thousand data points would be able to if 40 can already get us this much.

4. REFERENCES

1. Manthiram, Arumugam. "An outlook on lithium ion battery technology." *ACS central science* 3, no. 10 (2017): 1063-1069.
2. Liu, Kai, Yayuan Liu, Dingchang Lin, Allen Pei, and Yi Cui. "Materials for lithium-ion battery safety." *Science advances* 4, no. 6 (2018): eaas9820.
3. Wurzelmann, Sam. "Advanced Research Projects Agency-Energy (ARPA-E): Innovation through the US Department of Energy." Center for Climate and Energy Solutions (2012).
4. Cubuk, Ekin D., Austin D. Sendek, and Evan J. Reed. "Screening billions of candidates for solid lithium-ion conductors: A transfer learning approach for small data." *The Journal of chemical physics* 150, no. 21 (2019): 214701.
5. Sendek, Austin D., Qian Yang, Ekin D. Cubuk, Karel-Alexander N. Duerloo, Yi Cui, and Evan J. Reed. "Holistic computational structure screening of more than 12000 candidates for solid lithium-ion conductor materials." *Energy & Environmental Science* 10, no. 1 (2017): 306-320.

6. Sendek, Austin D., Ekin D. Cubuk, Evan R. Antoniuk, Gwoon Cheon, Yi Cui, and Evan J. Reed. "Machine learning-assisted discovery of many new solid li-ion conducting materials." arXiv preprint arXiv:1808.02470 (2018).
7. Sendek, Austin D., Ekin D. Cubuk, Evan R. Antoniuk, Gwoon Cheon, Yi Cui, and Evan J. Reed. "Machine learning-assisted discovery of solid Li-ion conducting materials." Chemistry of Materials 31, no. 2 (2018): 342-352.
8. Kamaya, Noriaki, Kenji Homma, Yuichiro Yamakawa, Masaaki Hirayama, Ryoji Kanno, Masao Yonemura, Takashi Kamiyama et al. "A lithium superionic conductor." Nature materials 10, no. 9 (2011): 682-686.
9. Jordan, Michael I., and Tom M. Mitchell. "Machine learning: Trends, perspectives, and prospects." Science 349, no. 6245 (2015): 255-260.
10. Ceder, G., and K. Persson. "The materials project: A materials genome approach." DOE Data Explorer, <http://www.osti.gov/dataexplorer/biblio/1077798>. [2016-08-28] (2010).
11. Huggins, Robert A. "Recent results on lithium ion conductors." In International Symposium on Solid Ionic and Ionic-Electronic Conductors, pp. 773-781. Pergamon, 1977.
12. Fujimura, Koji, Atsuto Seko, Yukinori Koyama, Akihide Kuwabara, Ippei Kishida, Kazuki Shitara, Craig AJ Fisher, Hiroki Moriwake, and Isao Tanaka. "Accelerated materials design of lithium superionic conductors based on first-principles calculations and machine learning algorithms." Advanced Energy Materials 3, no. 8 (2013): 980-985.
13. Meredig, Bryce, Ankit Agrawal, Scott Kirklin, James E. Saal, Jeff W. Doak, Alan Thompson, Kunpeng Zhang, Alok Choudhary, and Christopher Wolverton. "Combinatorial screening for new materials in unconstrained composition space with machine learning." Physical Review B 89, no. 9 (2014): 094104.
14. Dreiseitl, Stephan, and Lucila Ohno-Machado. "Logistic regression and artificial neural network classification models: a methodology review." Journal of biomedical informatics 35, no. 5-6 (2002): 352-359.
15. Vellido, Alfredo, José David Martín-Guerrero, and Paulo JG Lisboa. "Making machine learning models interpretable." In ESANN, vol. 12, pp. 163-172. 2012.

16. Ward, Logan, Ankit Agrawal, Alok Choudhary, and Christopher Wolverton. "A general-purpose machine learning framework for predicting properties of inorganic materials." *npj Computational Materials* 2, no. 1 (2016): 1-7.
17. Warner, John T. *The handbook of lithium-ion battery pack design: chemistry, components, types and terminology*. Elsevier, 2015.
18. <https://nptel.ac.in/courses/106/106/106106139/>
19. Wang, Shuo, Qiang Bai, Adelaide M. Nolan, Yunsheng Liu, Sheng Gong, Qiang Sun, and Yifei Mo. "Lithium Chlorides and Bromides as Promising Solid-State Chemistries for Fast Ion Conductors with Good Electrochemical Stability." *Angewandte Chemie International Edition* 58, no. 24 (2019): 8039-8043.



Plagiarism Checker X Originality Report

Similarity Found: 5%

Date: Wednesday, November 24, 2021

Statistics: 281 words Plagiarized / 6164 Total words

Remarks: Low Plagiarism Detected - Your Document needs Optional Improvement.

Abstract: Organic solvents are used as electrolytes in lithium-ion batteries currently. They work well by most accounts, lithium flows through them pretty quickly, they're manufactured at scale but organic solvents are a problem because they're very flammable.

Hence **there is a need** for entering the unknown part of material science where we are combining what we know from the conventional approaches with these new ideas from **big data and machine learning** to accelerate the current work of search for materials for solid electrolytes in the lithium-ion batteries. **The development of new solid electrolytes could** ease a lot of concerns over the safety, stability, energy density, and cycle life of commercial batteries.

Furthermore, **the development of new solid electrolytes could facilitate the development of structural batteries for the weight** and volume delicate applications of electric spacecraft and aircraft. Keywords: Solid Electrolyte, Machine Learning, Lithium-ion batteries, Superionic Conductors, Solid-State Batteries.