

Article

Data-driven studies of Li-ion-battery materials

Steven K. Kauwe ^{1,†}, Trevor David Rhone ^{2,†} and Taylor D. Sparks ^{1*}

¹ Materials Science & Engineering Department, University of Utah, Salt Lake City, UT, 84112

² The Department of Physics, Harvard University, Cambridge, MA, 02138

* Correspondence: sparks@eng.utah.edu; Tel.: +1-801-581-8632

† These authors contributed equally to this work.

Academic Editor: name

Version June 19, 2019 submitted to Crystals; Typeset by L^AT_EX using class file mdpi.cls

Abstract: Batteries are a critical component of modern society. The growing demand for new battery materials—coupled with a historically long materials development time—highlights the need for advances in battery materials development. Understanding battery systems has been frustratingly slow for the materials science community. In particular, the discovery of more abundant battery materials has been difficult. In this paper, we describe how machine learning tools can be exploited to predict the properties of battery materials. In particular, we report the challenges associated with a data-driven investigation of battery systems. Using a data set of cathode materials and various statistical models we predicted the specific discharge capacity at 25 cycles. We discuss the present limitations of this approach and propose a paradigm shift in the materials research process that would better allow data-driven approaches to excel in aiding the discovery of battery materials.

Keywords: battery materials; machine learning; materials discovery

1. Introduction

Battery materials research began with Volta's voltaic pile in the 1800's [1]. Since then, new materials have facilitated the development of new batteries; each with desired properties such as enhanced energy density, rechargeability, high power output, and low cost. Li-ion batteries were conceived in 1980 by Mizushima, Jones, Wiseman, and Goodenough [2]. Today they are one of the most popular batteries constituting approximately one third of all batteries [3] because of their high energy density and rechargeability.

Nevertheless, Li-ion batteries have a lot of room for improvement. For one thing, they are relatively high cost owing to their use of transition metals such as Co, Ni, and Mn. Additionally, reliance on organic electrolytes has lead to flammability and safety concerns upon dendrite growth. Another challenge is that high capacity anodes are still being developed or are not yet available. Therefore, we clearly see that each part of the battery would benefit from new materials being discovered, for example: an earth abundant cathode, a solid electrolyte, or a high capacity anode.

Currently, incremental local optimization is the primary tool for discovering new materials. However, it may be possible for data-driven methods to screen an enormous composition or micro-structural space. The principle being that machine learning (ML) could learn from data that exists in the literature, and then very rapidly predict properties of interest for an enormous number of battery material candidates. The use of machine learning as a tool for materials discovery is rapidly growing. Examples can be found in the fields of thermoelectrics [4–6], superhard materials [7], thermochemical data [8,9], electronic properties [10–14], structural materials [15], functional materials [16–18], and structure classification [19–23]. Given the history of success of ML methods, it's natural to want to apply them to battery materials research. There are $\sim 10^4$

35 battery publications per year ($> 7000/year$ for the last three years according to Scopus keyword
36 search). Thus, there may be ample data available to exploit using statistical models. Indeed, battery
37 researchers have recently shown an interest in employing machine learning for answering battery
38 research questions. For example, Hu *et al.* used machine learning to predict the state-of-charge in
39 batteries for electric vehicles [24]. Liu *et al.* showed that machine learning could successfully be
40 used to predict the remaining useful life for batteries systems [25]. In this paper, we apply ML
41 methods to a set of data describing Li-ion-battery architecture, properties, and performance. We
42 explore not only chemical descriptors, but also those associated with the architecture and battery
43 testing conditions. We observe several hurdles for the successful application of ML methods to battery
44 research, where heterogeneous materials and properties are common. We describe the challenges of
45 applying a machine learning approach for the discovery of battery materials [26–30].

46 2. Materials and Methods

47 The data used for this work were previously published by Ghadbeigi *et al.* as an analysis
48 article [31]. This work did not attempt any machine learning, but instead provided unique
49 visualization and sustainability metrics such as scarcity and Herfindahl-Hirschman index (HHI).

50 As a first step for machine learning, we needed to retrieve and clean the data. For the case of this
51 data set, there were extensive missing entries. This is largely due to poorly enforced standards for
52 battery data reporting. For example, one author might provide the cathode type including the mass
53 fraction of the binder and carbon additives, where another author might simply state that carbon
54 additive was used. As a result, potentially useful descriptors—such as rate performance, calcination
55 temperature, and synthesis route—need to be removed due to lack of useful information or excessive
56 missing entries. This greatly reduces the data available for analysis. Furthermore, we chose the
57 discharge capacity at 25 cycles as the target property for our analysis. Consequently, experimental
58 descriptors correlated with the discharge capacity at 25 cycles, such as the discharge capacity at
59 50 cycles, were also removed. After removing the problematic columns, if a row still had missing
60 entries it was dropped. Besides learning from the metadata, additional descriptors were generated
61 from the composition of the cathode material. These descriptors should be readily available (e.g.
62 from databases) or easily computed without the need for computationally expensive simulations.
63 Using the atomic properties of the constituent elements, we can construct a mathematical description,
64 or descriptor, of the composition. That is, we consider the weighted average, and the maximum
65 difference of a set of atomic properties for a given composition. An example of the atomic properties
66 include the ionic radius, electronegativity, and the number of valence electrons [32–34]. See the
67 Appendix for the entire list of descriptors.

68 Information regarding the microscopic and macroscopic structure of the cathode materials is
69 expected to play an important role in describing battery properties. However, since this information
70 was poorly reported in the literature, our data set did not possess structural descriptors beyond
71 identifying the cathode structure class. We believe that it is in the interest of the scientific community
72 that, in the future, information be standardized and more systematically reported in order to better
73 serve data-driven studies [35]. Reporting of failed results should also be given equal weight to that
74 of successful results.

75 The raw data were combined with additional data which were generated from creating
76 additional descriptors from atomic properties. The first step of the data analytics process involved
77 visualizing the data in such a way as to gain insights into their behavior. The data were then divided
78 into a training set and a test set. The initial training data had 215 entries and 49 descriptors. The
79 training data were used to perform k-fold cross validation. We exploited machine learning models to
80 approximate the relationship between descriptors of the battery materials and the target property, the
81 specific discharge capacity at 25 cycles (in mAh g^{-1}). We considered lasso, kernel ridge regression
82 (KRR), random forests (RF) and support vector regression (SVR) [36]. We used grid search to obtain
83 the hyper-parameters of our models in order to minimize the error on our test set. The random

84 forests model allowed us to calculate a list of descriptor importances, which describes the relative
 85 importance of the descriptors used in the model. The top 25 descriptors were selected and used in a
 86 second round of model training in order to reduce over fitting and improve prediction performance
 87 on the test set. See the appendix for a list of these descriptors. We report the test error for the models
 88 evaluated in terms of the mean absolute error and the coefficient of determination (R^2).

89 **3. Results and Discussion**

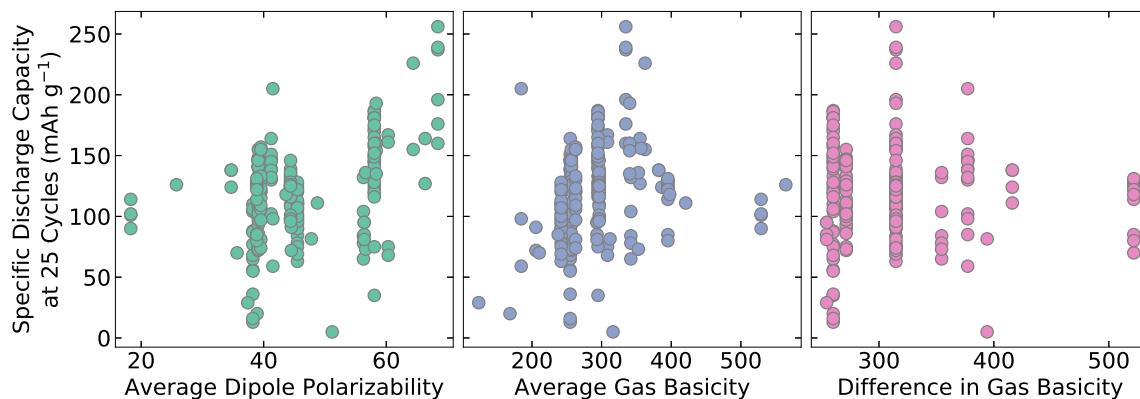


Figure 1. Specific discharge capacity at 25 cycle (mAh g^{-1}) versus the following descriptors: (a) mean of the dipole polarizability, (b) atomic volume maximum difference, and (c) the mean of the gas phase basicity.

90 Individuals interested in modeling and predicting battery performance care about a number of
 91 different metrics. In this work we use specific discharge capacity at 25 cycles as an example. We select
 92 the discharge capacity at 25 cycles because capacity fade is an important property which involves very
 93 time consuming testing, which we would rather predict via a high fidelity machine learning model.
 94 For example, if a battery is being tested at C/10 rate, then the discharge-charge cycle takes almost an
 95 entire day which means testing 25 cycles will take 20 days and testing 100 cycles would take nearly
 96 3 months. It is possible to visualize how this target property varies with the available descriptors in
 97 order to explore correlations [37] with other materials' properties. If obvious correlations exist, then
 98 a human can detect and exploit them using intuition. However, as can be seen in Figure 1 there is
 99 no obvious correlation. Nevertheless, machine learning algorithms are capable of building predictive
 100 models using higher-dimensional data not perceived by humans.

101 Lasso is a linear model, whereas KRR, SVR, and RF models allow for the learning of both linear
 102 and non-linear relationships. We attempt to map materials' descriptors to our target property using
 103 these models. Lasso, the only linear model, had the worst performance and is not reported here. The
 104 prediction accuracy of KRR, SVR, and RF models are shown in Figure 2. The model performance is
 105 below what is typically reported in the literature. This significant model error hinders the predictive
 106 utility for battery material discovery.

107 Descriptor importances can be extracted from the random forest model to determine which
 108 descriptors are most important for a successful prediction. Figure 2 shows the top three
 109 descriptors: average dipole polarizability, average gas phase basicity, and the difference in gas phase
 110 basicity. Electrochemical batteries rely on the insertion and removal of ions and the associated
 111 reduction-oxidation reaction. Therefore, the gas phase basicity (that is, the change of Gibbs
 112 free energy during deprotonation) is likely a pertinent materials descriptor which lends itself to
 113 mechanistic analysis. Other descriptors such as heat of fusion or dipole polarizability may play a
 114 more subtle or second-order role in determining Li-ion battery properties.

115 Although there are a number of battery materials with high values for the specific discharge
 116 capacity at 25 cycles, the vast majority of materials—all of which were investigated as potential

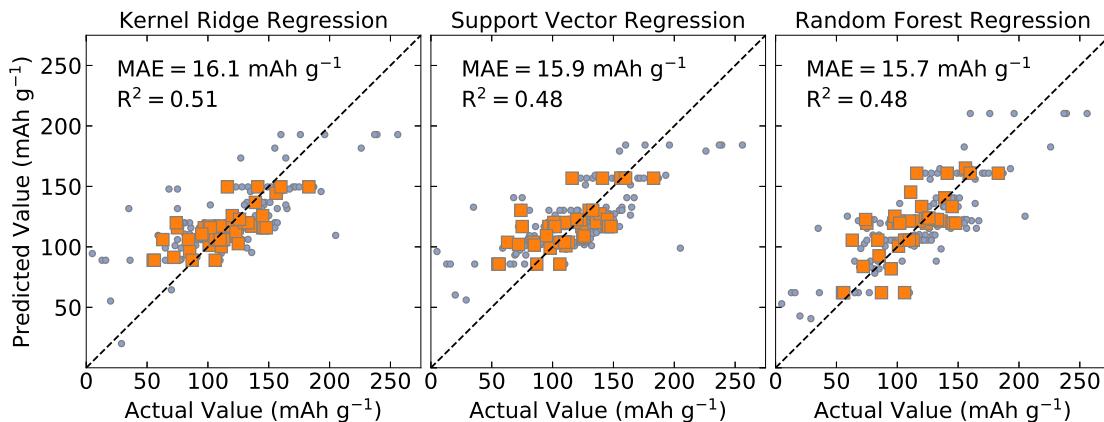


Figure 2. Machine learning prediction results for specific discharge capacity at 25 cycles (mAh g^{-1}) using (a) kernel ridge regression, (b) support vector regression and (c) random forests regression. Training (blue circles) and test (orange squares) performance show similar behavior. The mean absolute error (MAE) and coefficient of determination (R^2) are shown. Dotted lines are included to help visualize ideal model performance.

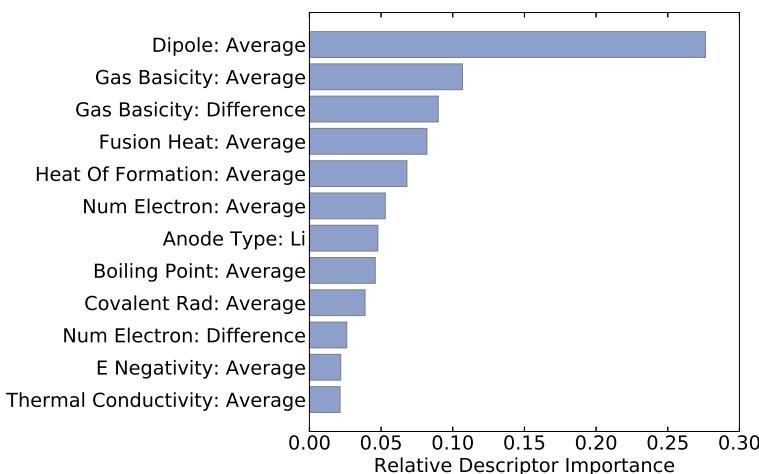


Figure 3. The top 12 descriptors are shown along with their importances from the random forest regression.

117 battery materials—have much lower values. A successful machine learning model would cheaply and
 118 accurately predict the properties of a battery system, allowing researchers to identify materials with
 119 desirable properties without the need for costly experiments or simulations. The model performance
 120 shown in the Figure 2 is not ideal, allowing for large prediction uncertainty. Although our model
 121 could conceivably distinguish between very small and large specific discharge capacity, the present
 122 predictive model would likely benefit from improvements in its performance prior to being used as
 123 a tool for materials discovery.

124 This lackluster model performance likely arises from several issues. An algorithm may not
 125 be capable of capturing underlying relationships, as is the case for lasso. Lasso is a linear model
 126 not capable of capturing nonlinear relationships in the data. In addition, the descriptors might be
 127 insufficient to adequately describe a battery system. The design of material descriptors plays a critical
 128 role in model performance and is an active area of research [32,38–41]. In particular, the complexity
 129 of battery space likely requires appropriate descriptors as well as additional data, for the successful
 130 training of ML models. The latter may be explored by studying how model error varies as a function
 131 of the size of the data set. In Figure 4, we see the error plateau at a relatively large value when the size

of the data set $N=90$. We find that the training error is significantly lower than the test error, indicating the impact of over-fitting. This suggests that the model error is dominated by the lack of appropriate descriptors. It is conceivable that some of the metadata columns which needed to be removed might be useful for improving model accuracy. This highlights a key challenge that distinguishes battery research from other fields. That is, the complexity of battery architecture and materials coupled with the lack of uniformity in data measurement and reporting hinders a data-driven study of these systems.

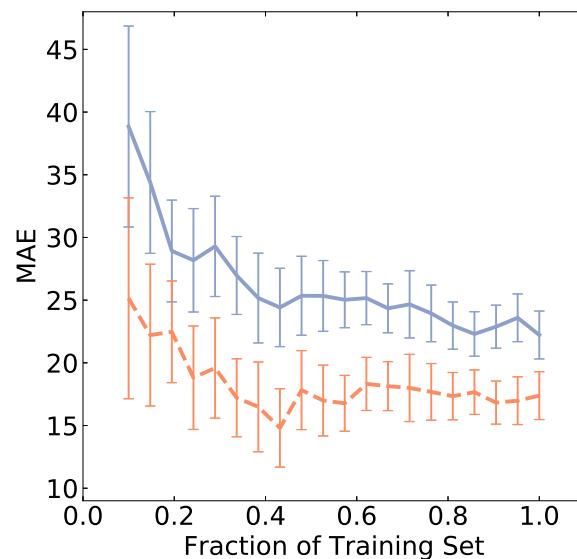


Figure 4. A decrease in mean absolute error (MAE) specific discharge capacity at 25 cycles (in mAh g^{-1}) is shown as a function of data set size using the SVR algorithm. Blue (orange) represents the performance of the test (training) set. Error bars show the standard deviation (obtained by repeating the calculation 50 times, each with a random sampling of the training data).

For example, consider the materials that constitute the battery components: anode, cathode, and electrolyte. We note that these are rarely single phase materials. Instead, a cathode is typically made up of the active material mixed together with conductive additives as well as a binder. The ratio of these phases is not dictated by an agreed upon standard but is determined on a case-by-case basis. The cathode can also retain significant porosity, or even gradients in porosity [42,43] which is not uniformly reported but will certainly influence battery properties. Likewise, the electrolyte in this data set is very often the organic LiPF_6 . This electrolyte can vary widely with different concentrations of additives, such as ethylene-carbonate, dimethyl carbonate, diethyl carbonate, and propylene carbonate. When it comes to assembling cells, even while testing cathode materials, researchers do not use a standard test cell architecture. In some cases the counter electrode could be lithium, in others graphite, and so forth.

In addition, we must consider the complexity which simply arises from non-standardized testing. For example, cells can be tested at different discharge rates, which can vary over several orders of magnitude. In other instances, researchers will report cyclic voltammograms which require a data scientist to interpret voltage ranges by hand when retrieving the data.

We summarize the key challenges that the community might consider addressing in order to increase the impact of a machine learning approach:

1. There are many cell architectures.
2. Each part of the cell is typically a composite, where the volume fraction, distribution, etc. are not always reported.
3. Standardized description of the crystal structure of the cathode material should be reported.

160 4. A uniform set of protocols for testing cells is not adhered to, and as a result reported data lack
161 compatibility.
162 5. The complexity associated with battery systems results in a large number of diverse descriptors.
163 Even if the data were reported in a uniform way, this large number of descriptors might require
164 large amounts of data for the accurate training of ML models.

165 **4. Conclusions**

166 The large amount of research being generated in battery materials provides an opportunity for
167 potentially advancing this area of research using data-driven tools, which will aid battery modeling
168 and the discovery of battery materials. This report highlights the challenges faced in generating
169 high fidelity models, and suggests a framework which will guide the battery community when
170 performing and reporting future experiments. Overcoming these challenges will benefit a data-driven
171 approach to battery research and serve the entire community. This calls for efforts to thwart the lack
172 of standardization which obfuscates experimental results and thus scientific advancement.

173 **Acknowledgments:** The authors would like to thank Efthimios Kaxiras and Leila Ghadbeigi for useful
174 discussions. We would also like to thank the National Science Foundation for their support of this research
175 under NSF CAREER Award 1651668.

176 **Author Contributions:** TDR and SKK performed data curation. TDR analyzed the data. SKK and TDS provided
177 domain guidance. SKK, TDS, and TDR wrote the paper.

178 **Conflicts of Interest:** The authors declare no conflict of interest.

179 **Abbreviations**

180 The following abbreviations are used in this manuscript:

181
182 ML: Machine learning
183 HHI: Herfindahl-Hirschman index
184 KRR: Kernel ridge regression
185 RF: Random forests
186 SVR: Support vector regression

187 **Appendix A. Descriptor and Model Details**

188 There were originally 49 descriptors that were considered (Table A1). The first two columns
189 of Table A1 show the element-based descriptors that were considered. The element values were
190 obtained using the Mendeleev package [44]. The far right column shows descriptors that were
191 obtained from the battery data. The number of descriptors was reduced to 25 using the descriptor
192 ranking from the random forest algorithm. All remaining descriptors used for training, along with
193 their importances, are shown in Table A2. The training/test data were obtained by using a split of
194 80/20. Model validation and parameter selection were performed using 10-fold cross-validation.

195 **Bibliography**

196 1. Finn, B. Origin of Electrical Power. *Essay from the National Museum of American History*,
197 <http://americanhistory.si.edu/csr/powering/prehist/prehist.htm> 2001.
198 2. Mizushima, K.; Jones, P.; Wiseman, P.; Goodenough, J.B. LixCoO₂ (0 < x < 1): A new cathode material for
199 batteries of high energy density. *Materials Research Bulletin* **1980**, *15*, 783–789.
200 3. Frost, S.; Sullivan. Technical insights report: Energy storage and generation for portable power. Technical
201 report, London: Frost Sullivan, 2009.
202 4. Sparks, T.D.; Gaulois, M.W.; Oliynyk, A.; Brigo, J.; Meredig, B. Data mining our way to the next
203 generation of thermoelectrics. *Scripta Materialia* **2016**, *111*, 10–15.

Table A1. The original 49 descriptors that were considered for learning. All averages are assumed to be weighted.

Element-based descriptor	Element-based descriptor	From battery data
Atomic Rad: Average	Ion: Average	Structure Type: Li3PO4
Atomic Rad: Difference	Boiling Point: Average	Structure: LiCdBO3
Atomic Vol: Average	Boiling Point: Difference	Structure: LiV3O8
Atomic Vol: Difference	Density: Average	Structure: Na2FePO4F
Covalent Rad: Average	Density: Difference	Structure: Nasicon
Covalent Rad: Difference	Evaporation Heat: Average	Structure: Fluorophosphate
Dipole: Average	Evaporation Heat: Difference	Structure: Layered
Dipole: Difference	Fusion Heat: Average	Structure: Multiple
Num Electron: Average	Fusion Heat: Difference	Structure: Olivine
Num Electron: Difference	Gas Basicity: Average	Structure: Spinel
Van der Waals Radius: Average	Gas Basicity: Difference	Anode: Li
Van der Waals Radius: Difference	Heat of Formation: Average	Anode: Li4Ti5O12
Electronegativity: Average	Heat of Formation: Difference	Anode: MCMB
Electronegativity: Difference	Melting Point: Average	Anode: Na
# Valence: Average	Melting Point: Difference	Anode: Graphite
# Valence: Difference	Thermal Conductivity: Average	
Ion: Average	Thermal Conductivity: Difference	

Table A2. The 25 descriptors used for reported model performance and their ranked importance.

Rank	Descriptor	Rank	Descriptor
1	Dipole: Average	14	Atomic Vol: Average
2	Gas Basicity: Average	15	Anode: Graphite
3	Gas Basicity: Difference	16	Electronegativity: Difference
4	Fusion Heat: Average	17	Melting Point: Average
5	Heat Of Formation: Average	18	Thermal Conductivity: Difference
6	# Electron: Average	19	Heat Of Formation: Difference
7	Anode: Li	20	Atomic Rad: Average
8	Boiling Point: Average	21	Van der Waals Radius: Average
9	Covalent Rad: Average	22	Melting Point: Difference
10	# Electron: Difference	23	Atomic Rad: Difference
11	Electronegativity: Average	24	Valence: Average
12	Thermal Conductivity: Average	25	Valence: Difference
13	Atomic Vol: Difference		

204 5. Gaultois, M.W.; Oliynyk, A.O.; Mar, A.; Sparks, T.D.; Mulholland, G.J.; Meredig, B. Perspective:
205 Web-based machine learning models for real-time screening of thermoelectric materials properties. *APL Materials* **2016**, *4*, 053213.

207 6. Oliynyk, A.O.; Sparks, T.D.; Gaultois, M.W.; Ghadbeigi, L.; Mar, A. Gd₁₂Co₅. 3Bi and Gd₁₂Co₅Bi,
208 Crystalline Doppelganger with Low Thermal Conductivities. *Inorganic chemistry* **2016**, *55*, 6625–6633.

209 7. Mansouri Tehrani, A.; Oliynyk, A.O.; Parry, M.; Rizvi, Z.; Couper, S.; Lin, F.; Miyagi, L.; Sparks, T.D.;
210 Brgoch, J. Machine learning directed search for ultraincompressible, superhard materials. *Journal of the
211 American Chemical Society* **2018**, *140*, 9844–9853.

212 8. Kauwe, S.K.; Graser, J.; Vazquez, A.; Sparks, T.D. Machine Learning Prediction of Heat Capacity for Solid
213 Inorganics. *Integrating Materials and Manufacturing Innovation* **2018**, pp. 1–9.

214 9. Carrasquilla, J.; Melko, R.G. Machine learning phases of matter. *Nature Physics* **2017**, *13*, 431.

215 10. Balachandran, P.V.; Kowalski, B.; Sehirlioglu, A.; Lookman, T. Experimental search for high-temperature
216 ferroelectric perovskites guided by two-step machine learning. *Nature communications* **2018**, *9*.

217 11. Zhuo, Y.; Mansouri Tehrani, A.; Brgoch, J. Predicting the Band Gaps of Inorganic Solids by Machine
218 Learning. *The journal of physical chemistry letters* **2018**, *9*, 1668–1673.

219 12. Kauwe, S.K.; Welker, T.; Sparks, T.D. Extracting Knowledge from DFT: Experimental Band Gap
220 Predictions Through Ensemble Learning **2018**.

221 13. Zhuo, Y.; Tehrani, A.M.; Oliynyk, A.O.; Duke, A.C.; Brgoch, J. Identifying an efficient, thermally robust
222 inorganic phosphor host via machine learning. *Nature communications* **2018**, *9*, 4377.

223 14. Hutchinson, M.L.; Antono, E.; Gibbons, B.M.; Paradiso, S.; Ling, J.; Meredig, B. Overcoming data scarcity
224 with transfer learning. *arXiv preprint arXiv:1711.05099* **2017**.

225 15. Meredig, B.; Agrawal, A.; Kirklin, S.; Saal, J.E.; Doak, J.; Thompson, A.; Zhang, K.; Choudhary, A.;
226 Wolverton, C. Combinatorial screening for new materials in unconstrained composition space with
227 machine learning. *Physical Review B* **2014**, *89*, 094104.

228 16. Cao, B.; Adutwum, L.A.; Oliynyk, A.O.; Luber, E.J.; Olsen, B.C.; Mar, A.; Buriak, J.M. How to optimize
229 materials and devices via design of experiments and machine learning: Demonstration using organic
230 photovoltaics. *ACS nano* **2018**, *12*, 7434–7444.

231 17. Kusne, A.G.; Gao, T.; Mehta, A.; Ke, L.; Nguyen, M.C.; Ho, K.M.; Antropov, V.; Wang, C.Z.; Kramer, M.J.;
232 Long, C.; others. On-the-fly machine-learning for high-throughput experiments: search for rare-earth-free
233 permanent magnets. *Scientific reports* **2014**, *4*, 6367.

234 18. Landrum, G.A.; Penzotti, J.E.; Putta, S. Machine-learning models for combinatorial catalyst discovery.
235 *Measurement Science and Technology* **2004**, *16*, 270.

236 19. Oliynyk, A.O.; Mar, A. Discovery of intermetallic compounds from traditional to machine-learning
237 approaches. *Accounts of chemical research* **2017**, *51*, 59–68.

238 20. Graser, J.; Kauwe, S.K.; Sparks, T.D. Machine learning and energy minimization approaches for crystal
239 structure predictions: A review and new horizons. *Chemistry of Materials* **2018**, *30*, 3601–3612.

240 21. Oliynyk, A.O.; Gaulois, M.W.; Hermus, M.; Morris, A.J.; Mar, A.; Bröggel, J. Searching for Missing Binary
241 Equiatomic Phases: Complex Crystal Chemistry in the Hf–In System. *Inorganic chemistry* **2018**.

242 22. Oliynyk, A.O.; Antono, E.; Sparks, T.D.; Ghadbeigi, L.; Gaulois, M.W.; Meredig, B.; Mar, A. High-throughput
243 machine-learning-driven synthesis of full-Heusler compounds. *Chemistry of Materials* **2016**, *28*, 7324–7331.

244 23. Oliynyk, A.O.; Adutwum, L.A.; Harynuk, J.J.; Mar, A. Classifying crystal structures of binary compounds
245 AB through cluster resolution feature selection and support vector machine analysis. *Chemistry of
246 Materials* **2016**, *28*, 6672–6681.

247 24. Hu, X.; Li, S.E.; Yang, Y. Advanced Machine Learning Approach for Lithium-Ion Battery State Estimation
248 in Electric Vehicles. *IEEE Transactions on Transportation Electrification* **2016**, *2*, 140–149.

249 25. Liu, D.; Zhou, J.; Pan, D.; Peng, Y.; Peng, X. Lithium-ion battery remaining useful life estimation
250 with an optimized Relevance Vector Machine algorithm with incremental learning. *Measurement* **2015**,
251 *63*, 143–151.

252 26. Ahmad, Z.; Xie, T.; Maheshwari, C.; Grossman, J.C.; Viswanathan, V. Machine Learning Enabled
253 Computational Screening of Inorganic Solid Electrolytes for Suppression of Dendrite Formation in
254 Lithium Metal Anodes. *ACS Central Science* **2018**, *4*, 996–1006.

255 27. Sendek, A.D.; Yang, Q.; Cubuk, E.D.; Duerloo, K.A.N.; Cui, Y.; Reed, E.J. Holistic computational
256 structure screening of more than 12000 candidates for solid lithium-ion conductor materials. *Energy
257 & Environmental Science* **2017**, *10*, 306–320.

258 28. Wu, B.; Han, S.; Shin, K.G.; Lu, W. Application of artificial neural networks in design of lithium-ion
259 batteries. *Journal of Power Sources* **2018**, *395*, 128–136.

260 29. Liu, D.; Zhou, J.; Pan, D.; Peng, Y.; Peng, X. Lithium-ion battery remaining useful life estimation
261 with an optimized Relevance Vector Machine algorithm with incremental learning. *Measurement* **2015**,
262 *63*, 143–151.

263 30. Hu, X.; Li, S.E.; Yang, Y. Advanced Machine Learning Approach for Lithium-Ion Battery State Estimation
264 in Electric Vehicles. *IEEE Transactions on Transportation Electrification* **2016**, *2*, 140–149.

265 31. Ghadbeigi, L.; Harada, J.K.; Lettiere, B.R.; Sparks, T.D. Performance and resource considerations of Li-ion
266 battery electrode materials. *Energy & Environmental Science* **2015**, *8*, 1640–1650.

267 32. Ghiringhelli, L.M.; Vyiral, J.; Levchenko, S.V.; Draxl, C.; Scheffler, M. Big Data of Materials Science:
268 Critical Role of the Descriptor. *Phys. Rev. Lett.* **2015**, *114*, 105503.

269 33. Seko, A.; Hayashi, H.; Nakayama, K.; Takahashi, A.; Tanaka, I. Representation of compounds for
270 machine-learning prediction of physical properties. *Phys. Rev. B* **2017**, *95*, 144110.

271 34. Rhone, T.D.; Chen, W.; Desai, S.; Yacoby, A.; Kaxiras, E. Data-driven studies of magnetic two-dimensional
272 materials **2018**. pp. 1–7, [\[1806.07989\]](#).

273

274 35. Ghiringhelli, L.M.; Carbogno, C.; Levchenko, S.; Mohamed, F.; Huhs, G.; Lüders, M.; Oliveira, M.; Scheffler, M. Towards efficient data exchange and sharing for big-data driven materials science: metadata and data formats. *npj Computational Materials* **2017**, *3*, 46.

275 36. Hastie, T.; Tibshirani, R.; Friedman, J.H. *The elements of statistical learning: data mining, inference, and prediction, 2nd Edition*; Springer series in statistics, Springer, 2009.

276 37. Reshef, D.N.; Reshef, Y.A.; Finucane, H.K.; Grossman, S.R.; McVean, G.; Turnbaugh, P.J.; Lander, E.S.; Mitzenmacher, M.; Sabeti, P.C. Detecting novel associations in large data sets. *Science* **2011**, *334*, 1518–1524.

277 38. Rupp, M.; Tkatchenko, A.; Müller, K.R.; von Lilienfeld, O.A. Fast and Accurate Modeling of Molecular Atomization Energies with Machine Learning. *Phys. Rev. Lett.* **2012**, *108*, 58301.

278 39. Hansen, K.; Biegler, F.; Ramakrishnan, R.; Pronobis, W.; von Lilienfeld, O.A.; Müller, K.R.; Tkatchenko, A. Machine Learning Predictions of Molecular Properties: Accurate Many-Body Potentials and Nonlocality in Chemical Space. *The Journal of Physical Chemistry Letters* **2015**, *6*, 2326–2331.

279 40. Riley, P. Tensor field networks: Rotation- and translation-equivariant neural networks for 3D point cloud. <https://arxiv.org/pdf/1802.08219.pdf>.

280 41. Thomas, N.; Smidt, T.; Kearnes, S.; Yang, L.; Li, L.; Kohlhoff, K.; Riley, P. Tensor field networks : Rotation- and translation-equivariant neural networks for 3D point clouds arXiv : 1802 . 08219v3 [cs . LG] 18 May 2018. [\[arXiv:1802.08219v3\]](https://arxiv.org/abs/1802.08219v3).

281 42. Al Khateeb, S.; Sparks, T.D. Spray pyrolysis of conductor-and binder-free porous FeS₂ films for high-performance lithium ion batteries. *Journal of Materials Science*, pp. 1–16.

282 43. Liu, L.; Guan, P.; Liu, C. Experimental and Simulation Investigations of Porosity Graded Cathodes in Mitigating Battery Degradation of High Voltage Lithium-Ion Batteries. *Journal of The Electrochemical Society* **2017**, *164*, A3163–A3173.

283 44. mendeleev – A Python resource for properties of chemical elements, ions and isotopes, ver. 0.4.3. <https://bitbucket.org/lukaszmentel/mendeleev>, 2014–.

299 © 2019 by the authors. Submitted to *Crystals* for possible open access publication under the terms and
300 conditions of the Creative Commons Attribution license (<http://creativecommons.org/licenses/by/4.0/>)