

High- T_c superconductor candidates proposed by machine learning

Siwoo Lee,^{1,2} Jason Hattrick-Simpers,^{2,3,*} Young-June Kim,^{4,†} and O. Anatole von Lilienfeld^{1,2,3,4,5,‡}

¹*Department of Chemistry, University of Toronto, St. George campus, Toronto, ON, Canada*

²*Acceleration Consortium, University of Toronto, St. George campus, Toronto, ON, Canada*

³*Department of Materials Science & Engineering,
University of Toronto, St. George campus, Toronto, ON, Canada*

⁴*Department of Physics, University of Toronto, St. George campus, Toronto, ON, Canada*

⁵*Vector Institute for Artificial Intelligence, Toronto, ON, Canada*

We cast the relation between chemical compositions of solid-state materials and their superconducting critical temperature (T_c) in terms of a statistical learning problem with reduced complexity. Training of query-aware similarity-based ridge regression models on experimental SuperCon data with (implicit) and without (ambient) high pressure entries achieves average T_c prediction errors of ± 10 K for unseen out-of-sample materials. Subsequent utilization of the approach to scan ~ 153 k materials in the Materials Project enables the ranking of candidates by T_c while taking into account thermodynamic stability and small band gap. Stable top three high- T_c candidate materials with large band gaps for implicit and ambient pressures are predicted to be $\text{Cs}_2\text{Sn}(\text{H}_2\text{N})_6$ (324 K), CsH_5N_2 (315K), $\text{Rb}_2\text{Sn}(\text{H}_2\text{N})_6$ (305 K), and $\text{H}_{15}\text{IrBr}_3\text{N}_5$ (189 K), $\text{H}_{12}\text{OsN}_5\text{Cl}_3\text{O}$ (161 K), $\text{B}_{10}\text{H}_{13}\text{I}$ (151 K), respectively. Stable top three high- T_c candidate materials with small band gaps for implicit and ambient pressures are predicted to be $\text{RbLiH}_{12}\text{Se}_3\text{N}_4$ (255 K), $\text{CeH}_{14}\text{Cl}_3\text{O}_7$ (246 K), $\text{Li}(\text{H}_3\text{N})_4$ (234 K), and $\text{ReH}_{30}\text{Ru}_2(\text{NCl})_{10}$ (127 K), $\text{AlH}_{18}\text{Ru}(\text{NF})_6$ (120 K), $\text{Sr}(\text{Li}_2\text{P})_2$ (117 K), respectively.

I. INTRODUCTION

A major unsolved question in the physical sciences is whether there exists at ambient pressure a superconductor with its superconducting critical temperature, T_c , at or above room temperature³. Under these conditions, such a material must exhibit the hallmark properties of superconductivity^{4,5}, *i.e.* the persistent absence of electrical resistance and the expulsion of magnetic fields (the Meissner effect) below a critical magnetic field strength (H_c). Its discovery would be of immense significance to various scientific and industrial applications, encompassing energy storage and electrical power transmission^{6–9}, nuclear fusion energy^{10–12}, computing^{13–16}, medicine^{17,18}, public transport^{19–21}, to particle physics^{22–25}. Unfortunately, the identification of candidates is difficult because the overwhelming majority of known superconductors have T_c 's near 0 K², making it challenging to learn from the properties of high- T_c samples and thereby propose novel candidates with higher T_c 's. While the Bardeen-Cooper-Schrieffer (BCS) theory^{26,27} can successfully explain superconductivity at low temperatures as a phonon-mediated process to form a Cooper pair, there currently exists no complete microscopic theory of high- T_c superconductivity²⁸. Therefore, approaches to designing new higher- T_c materials have historically been largely empirical without the guidance of theory²⁹. While *ab initio* electronic structures methods could be more efficient

than experiments, the lack of predictive T_c modeling also impedes the computational design and discovery. This poses a severe bottleneck on the search for new superconductors within the vast materials compound space estimated by some to be populated by $\sim 10^{100}$ plausible materials³⁰.

In this work, we describe a similarity-based machine learning (ML) approach that is capable of extrapolating beyond the distribution of experimentally-measured T_c values found in the SuperCon data set² to potentially identify materials with T_c 's greater than the currently known highest value, which is 250 K for LaH_{10} ³¹. Recently, there have been many efforts to predict T_c 's directly from the crystal structures and chemical compositions of materials using ML^{32–39}. However, as far as we are aware, none of these works has successfully addressed the out-of-domain (OOD) problem^{40,41} that is inherent to attempts of making accurate predictions for samples with label values found beyond the range that the ML models were trained on. This is, in fact, a well-known issue in the ML field^{42,43}, and it is imperative that it is tackled if ML models are to be used to propose novel high- T_c candidates.

We have tackled this problem using the procedure outlined in Figure 1. First, for a given material, its n -nearest neighbors in the SuperCon^{2,44} training data sets are queried. These n -nearest neighbors are then used to train a ridge regression model, from which the test sample's T_c is predicted. Leave-one-out prediction tests of the resulting models trained on sets drawn either from all of the ~ 13.6 k samples in SuperCon² (implicit model), or from its subset exclusively containing entries at ambient pressures (ambient model), suggest

* jason.hattrick.simpers@utoronto.ca

† youngjune.kim@utoronto.ca

‡ anatole.vonlilienfeld@utoronto.ca

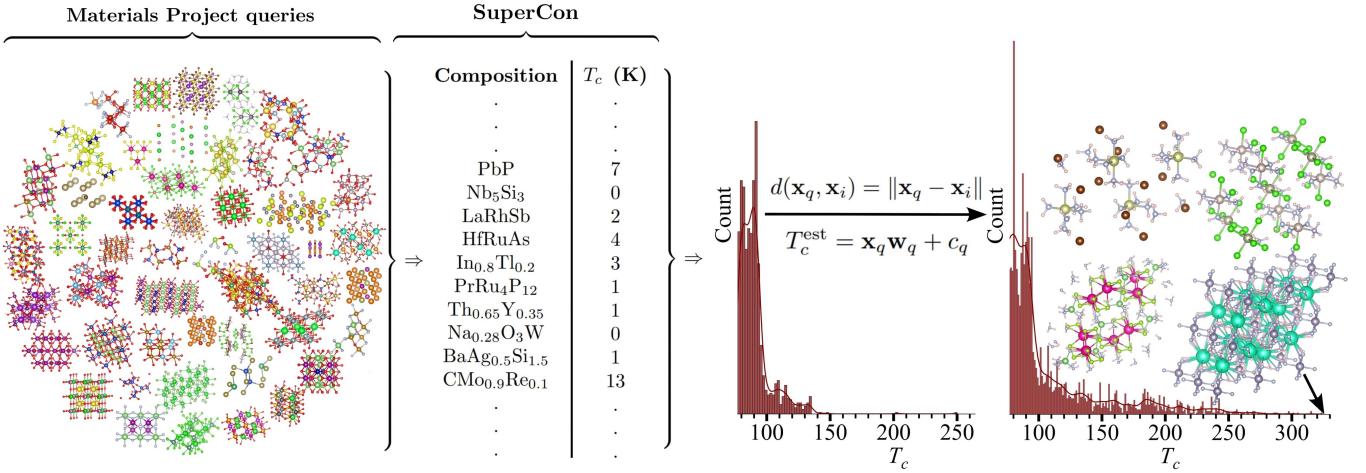


FIG. 1. Superconductor discovery using our similarity based ML methodology. From left to right: Query candidates (structural depictions drawn from the Materials Project¹) are used to select closest instances (in terms of d) from the SuperCon data set² with known experimentally-measured T_c 's for training ridge regression models. Materials representations do not require any structural information and are based only on features derived from chemical compositions. The application of this similarity-based machine learning to ~ 153 k materials in the Materials Project enriches the T_c distribution and includes multiple promising high- T_c superconductors not present in SuperCon.

good predictive performance throughout the full range of measured T_c 's. Subsequently, we have applied this approach to predict the T_c 's of the ~ 153 k materials for which calculated density functional theory (DFT) property results are recorded in the Materials Project¹. After filtering for thermodynamic stability, we have ranked all remaining materials by our ML T_c estimates. The implicit and ambient pressure model identifies thirty-five and six candidate materials exhibiting predicted T_c 's above 250 K and 135 K (the current record at ambient pressures⁴⁵), respectively. Filtering the ranked list also for DFT band gaps smaller than 1 eV, the top ranking materials of the implicit and ambient pressure models exhibit predicted T_c 's of 255 K and 127 K, respectively.

II. METHODOLOGY

A. Superconductors Data Set

In this work, the SuperCon data set² was obtained from the Materials Data Repository, which is maintained by the Japanese National Institute for Materials Science (NIMS). During the time which this work was conducted, SuperCon listed 26,321 materials, each with their associated chemical compositions, experimentally-measured T_c 's (Kelvin), and article references. It does not include atomic coordinates required to construct unit cells and does not mention whether the T_c measurements were performed while external pressures were applied to the samples. We cleaned the data set by assigning to stoichiometries with multiple T_c measurements their mean values.

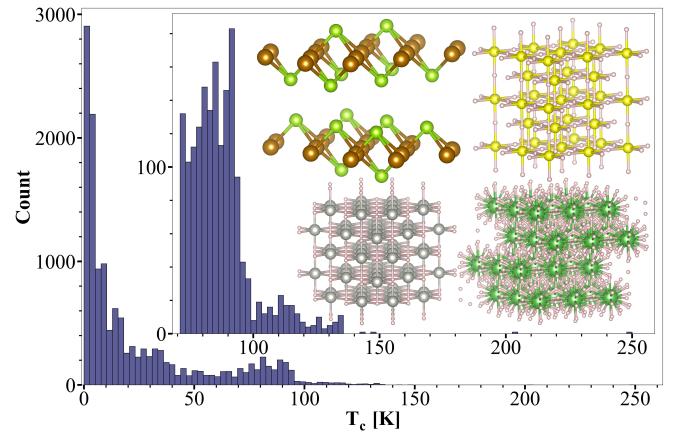


FIG. 2. Distribution of 13,661 T_c measurements in the curated SuperCon data set, including samples with T_c 's measured under ambient and applied pressures. Inset shows the distribution of samples with T_c 's measured to be greater than 70 K and four exemplary crystal structures.

Some contentious high- T_c measurements^{46–48} were also removed. As it was deemed to be prohibitively time-consuming to verify via human-effort the repute and accuracy of each measurement in SuperCon, it was assumed that the issue of the inclusion of measurements from untrustworthy experiments could be largely neglected in the low- T_c regime where the overwhelming majority of samples lie. Further, samples consisting of one or ten different element types were removed, as were those specified by arbitrary doping concentrations (e.g. Hg₂Ba₂YCu₂O_{8-x}).

The final, cleaned data set includes 13,661 unique

stoichiometries with T_c 's ranging 0–250 K, with a median value of 10 K (Figure 2). Using the definition of high-temperature superconductors as those with T_c 's greater than the boiling point of liquid nitrogen, 77 K⁴⁹, there exists only 1,372 such samples in the data set, constituting 10 % of its total size – clearly, the distribution is highly right-skewed. To illustrate, the four samples in the data set with the highest T_c 's are LaH₁₀, H₂S, H₃S, Hg_{0.66}Pb_{0.34}Ba₂Ca_{1.98}Cu_{2.9}O_{8.4}, with corresponding T_c 's of 250³¹, 203⁵⁰, 147⁵¹, and 143⁵² K, respectively. It may be of interest to note that experimental studies suggest the superconducting phase of H₂S which occurs under high pressures (gigapascals) actually exists as H₃S due to decomposition⁵⁰. While this mechanism may make redundant the inclusion in the data set of both H₂S and H₃S, they were retained for chemical diversity in this low-data regime.

The SuperCon² data set⁴⁴ – which was compiled by autonomously scraping the literature to extract data related to superconductors – was then used to determine the samples in SuperCon with T_c measurements performed under applied pressure. This analysis suggests that thirty-seven samples were measured under applied pressures. These were removed from SuperCon to create a separate data set containing only samples with T_c 's measured under ambient conditions (0–135 K).

B. Similarity-Based Machine Learning

To estimate the T_c of a given test sample via similarity-based ML⁵³, the training data set is first queried to find its n -nearest neighbors. The distance metric we employ is the Euclidean distance, calculated as

$$d(\mathbf{x}_q, \mathbf{x}_t) = \|\mathbf{x}_q - \mathbf{x}_t\| \quad (1)$$

where \mathbf{x}_q is the feature vector of the query test sample and \mathbf{x}_t is the feature vector of t 'th training set sample.

The n training samples with the smallest norm values are then chosen to train a ridge regression model with the optimal α hyperparameter⁵⁴ selected as the one yielding the best performance on the training set (smallest mean squared error), evaluated by n -fold cross-validation. A prediction for a given test sample is made as

$$T_c^{\text{est}} = \mathbf{x}_q \mathbf{w}_q + c_q \quad (2)$$

where c_q is the intercept and \mathbf{w}_q is the vector of weight coefficients obtained from the training set as the closed-form solution of

$$\mathbf{w}_q = (\mathbf{X} \mathbf{X}^\top + \alpha \mathbf{I})^{-1} \mathbf{X}^\top \mathbf{y} \quad (3)$$

for \mathbf{X} the n -sample feature matrix and \mathbf{y} the corresponding n -dimensional labels vector (T_c).

We perform all our computations with the scikit-learn^{55–57} library in Python and perform the matrix inversion using the Cholesky decomposition⁵⁴. The absolute values of predictions are used for the final estimates to avoid non-physical negative T_c values.

C. Machine Learning Representations

Since only chemical compositions are provided in SuperCon, they form the basis of the representations used to train our ML models. Despite their apparent simplicity, chemical compositions have been shown to be sufficient in accurately learning various properties of materials^{58–62}. They are also less restrictive in suggesting interesting stoichiometries to experimentalists, as the predictions derived from them are not specific to particular crystal structures contained in their convex hulls⁶³.

147 features were generated for each sample from its composition using the materials informatics Python library Matminer⁶⁴. The minimum, maximum, range, mean, average deviation, and mode were calculated from the following atomic properties, with weights given by stoichiometric coefficients: atomic number; Mendeleev number; atomic weight; melting temperature of elemental solids; periodic table group number; periodic table period number; covalent radius; electronegativity; number of filled s , p , d , f orbitals; number of valence electrons; number of unfilled s , p , d , f orbitals; number of unfilled valence orbitals; volume of elemental solid; band gap of elemental solid; magnetic moment of elemental solids; and space group number of elemental solids. Matminer was used to additionally calculate transition metal fractions; stoichiometric 0-, 2-, 3-, 5-, 7-, and 10-norms; average number of valence electrons in each of s , p , d , f orbitals; and the fractions of valence electrons in each of s , p , d , f orbitals.

D. Materials Project

We apply our similarity-based ML method to ~153k samples listed in the Materials Project database to screen for potential novel superconductors with high- T_c 's. For each material represented as described in section II C, predictions are made at implicit/ambient pressure by selecting n training samples from the cleaned implicit/ambient pressure SuperCon data set and training a ridge regression model, as described in section II B.

III. RESULTS AND DISCUSSION

A. SuperCon based validation of our approach

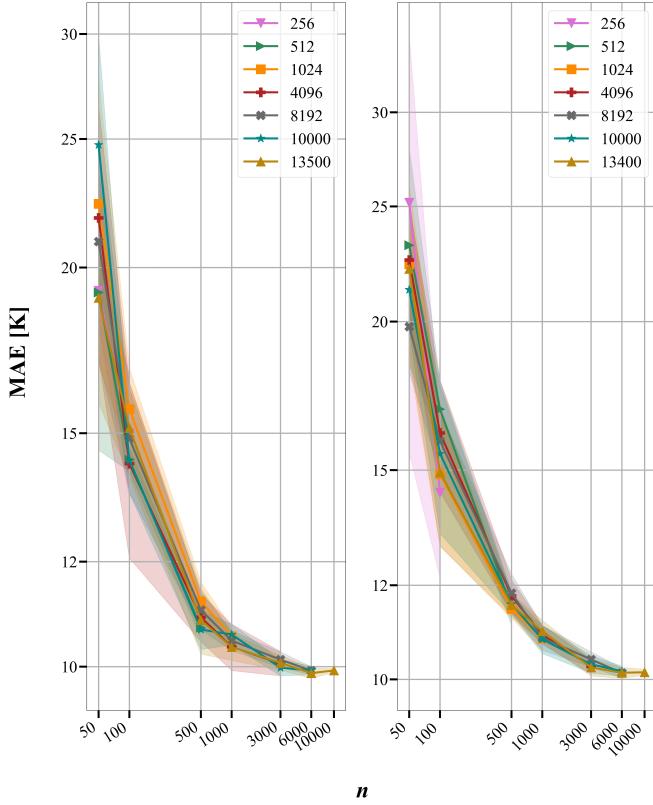


FIG. 3. Implicit pressure (left) and ambient pressure (right) model learning curves. Prediction errors for T_c 's of unseen out-of-sample test materials vs. training on n nearest neighbors selected from variable training pool sizes (indicated in the legend). The shading quantifies standard deviation in the predictions of a fixed test set at each n , obtained by five random selections of the training pool from the full data set.

The predictive performances of similarity-based ridge regression models in mapping chemical compositions to their superconducting T_c 's in SuperCon are first assessed with learning curves⁶⁵. Due to their characteristic shape, learning curves are useful for evaluating not only the data-efficiency of ML models but also the learnability of the problem. In Figure 3, the learning curves show the prediction error (mean absolute error, MAE) on the test set after training as a function of n nearest neighbors. Different learning curves are plotted where n training samples are selected from different pool-sizes of the full implicit/ambient pressure data set. The shadings indicate the standard deviations in the predictions obtained by training five different models on n -nearest neighbors selected from five random pools of the full data set. For example, in Figure 3, one curve is labelled “256” to indicate that 256 samples were randomly designated as the training data set from the

full implicit/ambient pressure data set. From this pool of 256 samples, the model performances after training on different sized subsets, for $n \leq 256$, are evaluated on the hold-out test set. The random pool selection and the training on n samples is performed five times at each point of the learning curves. Note that the hold-out test set selected from the implicit/ambient pressure data set was held constant throughout the generation of all the learning curves.

It is clear from each learning curve that increasing n produces systematic decay in the MAE's of the predicted T_c 's. However, contrary to noise-free labels and representations, the performance of our model saturates at an error of only ~ 10 K after training on 3000 samples for the implicit pressure models and after training on 6000 samples for the ambient pressure models. While this residual error is similar or only slightly larger than prediction errors of T_c reported previously by others^{32–38}, we stress again that our models, by contrast, are not incapable of extrapolation. The deviation from the expected linear decrease in the error with n ⁶⁵ suggests that either the choice of composition-based representations to train the ML models is not sufficiently unique for each sample, or that the labels are noisy, or both. It might also imply that the complex (non-linear) mapping from the features space to the labels space cannot be fully learned with linear predictor functions. However, some studies also suggest that the performance saturation may occur without regards to the selection of the learning algorithm. For instance, the use of kernel ridge regression in similarity-based learning of the atomization energies of molecules also exhibits the levelling in prediction errors⁵³. Nonetheless, our similarity-based ML method reaches prediction errors on the order of ~ 10 K, which, in the context of the large range of experimentally-measured T_c values spanned by different materials in SuperCon, we assume in the following to be sufficiently accurate and transferable for obtaining robust rankings among high- T_c candidate materials.

With the similarity-based ML approach, leave-one-out T_c predictions under unknown/ambient pressure are made for each of the 13,661/13,624 materials in the implicit/ambient pressure data set. The implicit pressure prediction for each sample was made by training a model on its 3000 nearest neighbors in the implicit pressure data set. $n = 3000$ was selected since, as can be seen from Figure 3, it already reaches an MAE of ~ 10 K. Hence, we can avoid the increased computational cost of the matrix inversion involved in ridge regression that would be incurred from training on a larger number of samples. For the same reasoning we selected $n = 6000$ for the ambient pressure models (See Fig. Figure 3 (Right)). Note that overall, however, the execution of this method is extremely efficient due to the relatively small size of n and the low cost associated to evaluate

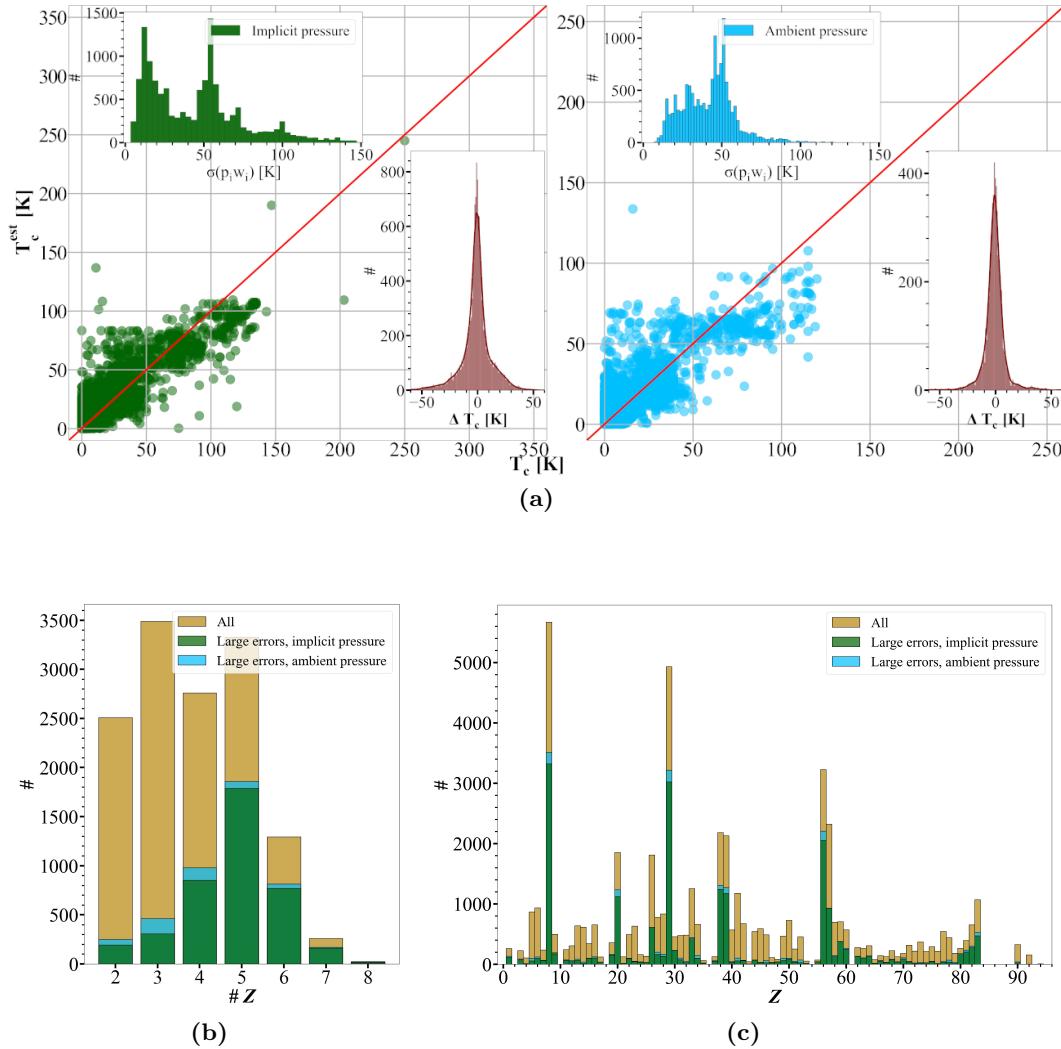


FIG. 4. (a) Leave-one-out T_c predictions under implicit/ambient pressure for 7,629/6,183 samples in the implicit/ambient pressure SuperCon data set with spreads in their index-wise feature-weight products of less than 51/42 K, achieving MAE $\sim 5/6$ K and $R^2 \sim 0.79/0.69$. Insets display these samples’ distributions of index-wise feature-weight products and signed-errors. (b) Distributions of the number of different elements and (c) and element types in the compositions of the SuperCon samples with absolute prediction errors greater than 10 K (without filtering for spreads in feature-weight products), compared against all SuperCon samples.

inner products as similarity measures. To illustrate, the total wall time required to search for n -nearest neighbors, tune for the α hyperparameter, train on the n training samples, and predict for the test sample is about a second on a laptop (12th gen. Intel i7-1260P, 12 cores, 2.10 GHz clock rate). Consequently, scanning a materials library for T_c requires 1 s per material and node.

Upon examination of the distribution of the index-wise products in the dot product between a query sample’s feature vector, \mathbf{x}_q , and the vector of weight coefficients, \mathbf{w}_q

$$\mathbf{x}_q \mathbf{w}_q = \sum_{i=1}^{147} x_{q,i} w_{q,i} \quad (4)$$

it emerges that some samples have large standard deviations in their product values (Figure 4a). This suggests that the α parameter, obtained from the training set, is not generalizable to the test sample as it fails to penalize the large coefficients that produce large $|x_{q,i} w_{q,i}|$ values. The removal of such implicit/ambient pressure samples with large spreads, defined as those greater than the mean spread of 51/42 K, improves predictive performances in terms of leave-one-out predictions on SuperCon (Figure 4a), as the MAE can be reduced from $\sim 9/10$ K to $\sim 5/6$ K. Although there are few samples with large prediction errors, similarity-based ML appears to achieve relatively good performances across the full range of T_c values. For instance, LaH₁₀ with T_c of 250 K is accurately predicted and no value greater than $\sim 250/135$ K is predicted

for the implicit/ambient pressure SuperCon materials, which corroborates the robustness of the models. The signed-errors distribution is symmetrical but not normally distributed. The former suggests that the models' predictions do not carry a systematic error. The non-normal distribution with large tails is due to the unphysical statistical nature of the outliers (particularly prominent in the scarce data regime the similarity-based ML model is operating in), and has previously already been observed and discussed⁶⁶.

An examination of the samples with absolute prediction errors greater than 10 K reveals they mostly comprise of compositions consisting of five different elements and that the error-rate proportion increases with greater number of differing elements in the system (Figure 4b). This may not be unexpected since the different number of environments an element can experience in its crystal system increases with increasing number of differing elements. However, in SuperCon, the number of samples as a function of differing elements does not appear to increase sufficiently for the different environments to be adequately accounted for in our ML models. We also note that large errors are associated with materials consisting of oxygen, copper, or barium (Figure 4c). This may be the effect of a combination of a bias incurred from intense research by the superconductor community into yttrium barium copper oxide and related compounds, and that oxygen as a strong oxidant complicates the chemistry involved in these systems.

The weights obtained from the training of ridge regression models were also inspected to better understand which features made the most significant contributions to the predictions of T_c labels. On average for the samples in the implicit pressure data set, the five most significant features, in decreasing order of importance as calculated by the magnitude of the absolute value of the product of the weight with its corresponding feature value $|x_{q,i}w_{q,i}|$, are the average atomic number, average covalent radius, average atomic weight, average deviation of atomic numbers, and average deviation of atomic weights. Correspondingly, in the ambient pressure data set, the five most important features are the average periodic table period number, average atomic number, average atomic weight, average deviation of atomic numbers, and average deviation of atomic weights. Clearly, the T_c predictions are greatly influenced by the masses and sizes of the elemental species in a given material's stoichiometry. These results are logical considering the observations of the isotope effect in both conventional and unconventional superconductors^{67–70}, in which their T_c 's are inversely proportional to the square-root of the atomic masses of the isotopes in their compositions as lighter ones produce higher phonon frequencies.

B. Rediscovery of Known Superconductors

We further evaluated our ML method by predicting T_c 's of published superconducting materials that are *not* part of the SuperCon data sets. Specifically, the two compounds $\text{Ni}(\text{TePd})_2$ ³⁹/(SrCa)₁₀Cu₁₇O₂₉⁷¹ were experimentally-measured in 2023/2000 to exhibit T_c 's of $\sim 1/75$ K. Our ambient pressure model predicts T_c for these samples to correspond to 13 K and 68 K, respectively, which are very close or even within the ~ 10 K error expected from our learning curves (Figure 3) and leave-one-out predictions (Figure 4a). These results represent an entirely independent test set and further corroborate our assumption that the approach is sufficiently robust for the identification of high- T_c material candidates.

C. Application to the materials project

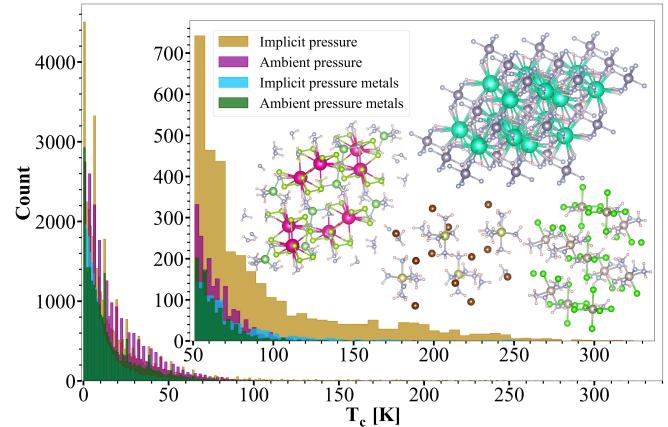


FIG. 5. Distribution of implicit/ambient pressure T_c predictions for Materials Project samples with energies above convex hulls of less than 0.030 eV/atom (38,539/27,821 samples), with subsets filtered for band gaps of less than 1.000 eV (23,402/25,186 samples). Inset shows the distributions of samples with predicted T_c 's greater than 50 K and representative crystal structures.

Our similarity-based ML approach has been used to estimate T_c 's of ~ 153 k materials listed in the Materials Project. The Materials Project was queried because it provides DFT-computed estimates of various materials properties, such as stability and band gap. Due to the computational efficiency of our approach, the entire scan consumed only ~ 42 node hours. We note that due to the generality of our approach, *any* other materials library (e.g. with experimental stability and band gap values) could have been used just as well.

Each sample's predictions under implicit pressure ($n = 3000$) and ambient pressure ($n = 6000$) were made by training on its nearest neighbors found in the implicit

and ambient pressure SuperCon data sets, respectively. Samples with implicit and ambient pressure predictions associated with standard deviations larger than 51 K and 42 K, respectively, are disregarded. Those with computed energies above their convex hulls of greater than 0.030 eV/atom are also disregarded as being thermodynamically unstable.

TABLE I. Compositions of materials in the Materials Project with the highest predicted T_c 's under implicit pressures, that lie < 0.030 ev/atom above their convex hulls.

ID	Composition	T_c^{est} (K)
mp-505233	$\text{Cs}_2\text{Sn}(\text{H}_2\text{N})_6$	324
mp-1184037, mp-1198227	CsH_5N_2	315
mp-643359	$\text{Rb}_2\text{Sn}(\text{H}_2\text{N})_6$	305
mp-643371	$\text{K}_2\text{Sn}(\text{H}_2\text{N})_6$	301
mp-643158	PH_8IN_4	297
mp-1193046	$\text{CsMgAs}(\text{H}_6\text{O}_5)_2$	297
mp-974267	KH_8N_3	288
mp-721084	H_7IN_2	287
mp-1202629	$\text{CdH}_{20}\text{N}_6\text{OF}_2$	286

TABLE II. Compositions of materials in the Materials Project with the highest predicted T_c 's under ambient pressure, that lie < 0.030 ev/atom above their convex hulls.

ID	Composition	T_c^{est} (K)
mp-1205028	$\text{H}_{15}\text{IrBr}_3\text{N}_5$	189
mp-24461	$\text{H}_{12}\text{OsN}_5\text{Cl}_3\text{O}$	161
mp-1199051, mp-1199255	$\text{B}_{10}\text{H}_{13}\text{I}$	151
mp-30977	$\text{B}_5\text{H}_6\text{Br}$	149
mp-1197561	$\text{B}_{10}\text{H}_{13}\text{Br}$	143
mp-634446	$\text{CsAl}(\text{H}_2\text{N})_4$	132
mp-1199374	$\text{TaH}_9\text{N}_3\text{F}_5$	132
mp-1204398	$\text{CoMoH}_{24}\text{N}_6\text{ClO}_7$	129
mp-767240	$\text{CsPH}_4(\text{NO})_2$	129

The resulting distributions in T_c predictions is shown in Figure 5. This analysis suggests that at unknown, potentially very high pressures, thirty-five materials may have T_c 's greater than 250 K. Of these materials, fourteen may transition to their superconducting phase at temperatures greater than 273 K (see Table I for representative samples), with the highest T_c predicted to be 324 K for $\text{Cs}_2\text{Sn}(\text{H}_2\text{N})_6$ ^{72,73} (Figure 6a). When filtering for metal-like materials with small band gaps (less than 1.000 eV), only one sample of $\text{RbLiH}_{12}\text{Se}_3\text{N}_4$ is predicted to have a T_c (255 K) greater than 250 K. At ambient pressure, six materials are predicted to have T_c 's greater than 135 K (143–189 K) (see Table II), while metal-like materials are predicted to reach only up to 127 K.

It is believed that one of the identifying characteristics

of high- T_c superconductors are strongly correlated bands that allow for unorthodox Cooper pair formations^{74,75}. For instance, from Figure 6b, it appears that $\text{Cs}_2\text{Sn}(\text{H}_2\text{N})_6$ has a pair of bands at its Fermi level that are relatively flat and narrow with a maximum bandwidth of \sim 160 meV. By the assumptions about band flatness correlating with superconductivity, this result may suggest that this material can transition to its superconducting phase at higher temperatures. However, band structure calculations for any given material often involve severe approximations and produce results that can deviate significantly (usually as underestimates) from experimental results^{76,77}. Therefore, it would be of great interest to verify the superconducting properties and electronic band structures of $\text{Cs}_2\text{Sn}(\text{H}_2\text{N})_6$ via experimental synthesis and measurements. Moreover, it would be interesting to conduct experiments of the thermodynamic stability of $\text{Cs}_2\text{Sn}(\text{H}_2\text{N})_6$ because the calculations conducted by the Materials Project state that its crystal structure is not the most stable configuration as it lies \sim 0.017 ev/atom above this stoichiometry's convex hull^{72,73}. After all, it would be of very limited practical usefulness if its superconducting phase occurs only under extremely high pressures. It appears that there currently has been no efforts to conduct such experiments, as this material is not reported in SuperCon and, to the best of our knowledge, it has not been studied in the literature.

IV. CONCLUSION

We have introduced a data-efficient similarity-based ML approach to estimating the superconducting critical temperatures of materials at both implicit and ambient pressures. Predictions for novel materials require training a query aware ridge regression model 'on the fly' using only the n nearest neighbours in the training data. This is feasible thanks to the extremely low computational cost required by our model (\sim 1 second/material on a twelve-core CPU). When this simple method was evaluated via leave-one-out predictions on the full SuperCon data set, it was found to be relatively robust in making accurate predictions across the full range of T_c values, 0–250 K. Moreover, it was observed that one may be able to identify predictions with large uncertainties as samples with large standard deviations in their index-wise products contained in their feature-weights vectors dot products. The analysis of the weight coefficients revealed that material properties related to the atomic weights and radii of the elements in the stoichiometries were the most significant contributors to the predictions of T_c 's, suggesting that the ML models were able to at least capture the basic physical principles involved in superconductivity.

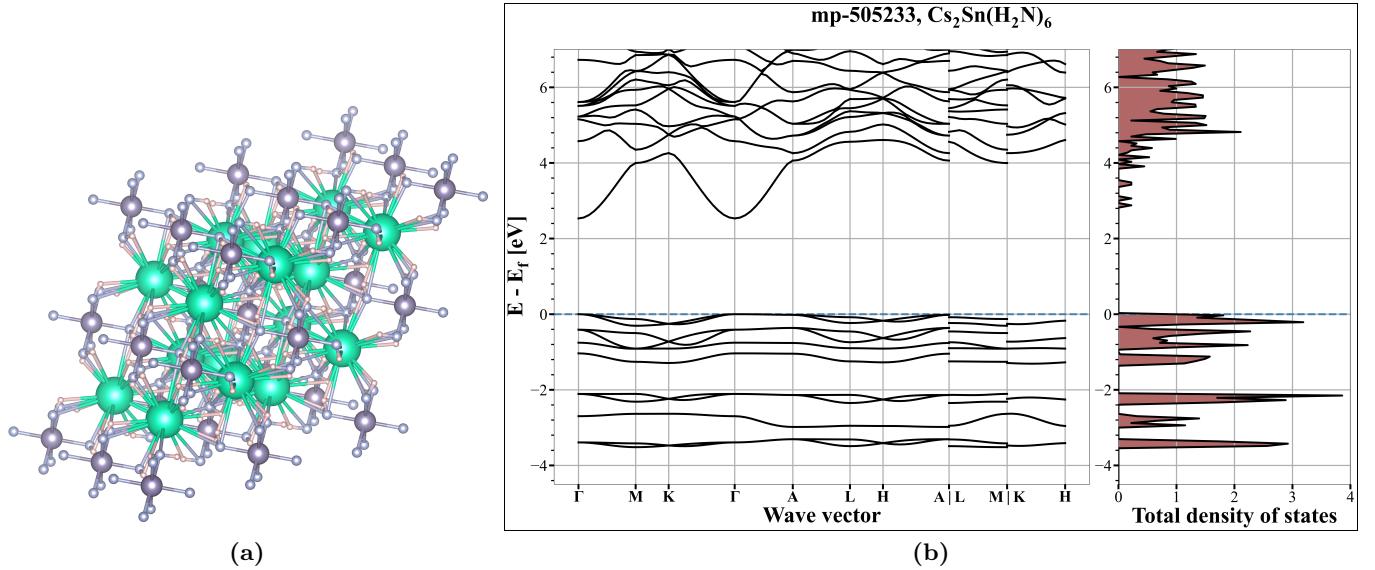


FIG. 6. Crystal structure and band structure of material in Table I with highest predicted T_c . (a) Crystal structure of $\text{Cs}_2\text{Sn}(\text{H}_2\text{N})_6$, and (b) its electronic band structures and density of states. The Fermi level is set to 0 eV, indicated by the dashed line. All data downloaded from materials project^{72,73}.

We have used the model to rank the entire Materials Project data base of $\sim 153k$ materials by the estimated T_c . Several materials were identified as potentially being able to exist as superconductors near room-temperature, albeit under unknown pressure, with $\text{Cs}_2\text{Sn}(\text{H}_2\text{N})_6$, despite its large band gap (~ 2.5 eV according to DFT), being of particular interest with an estimated T_c of 324 K at a pressure that would remain to be determined. Its electronic band structures records in the Materials Project indicate, however, that there is a flat band near its Fermi level, which possibly could be leveraged upon doping. As superconductivity experiments appear to have not yet been performed for $\text{Cs}_2\text{Sn}(\text{H}_2\text{N})_6$, nor for the other high- T_c materials identified, they might be valuable targets for future research.

Our results also exemplify the usefulness of similarity-based ML in accelerating the virtual design and discovery of novel materials and molecules with interesting properties. We note this may be facilitated by the specific choice of a representation which is more or less agnostic (depending on use case). For example, as done in this study, features only derived from chemical composition may be advantageous in allowing for the screening of a much greater number of materials as it would be trivial to create new stoichiometries by combinatorics of element types and stoichiometric coefficients. Further important criteria, such as measures of synthesizability or stability, could be added as constraints later on.

V. DATA AND CODE AVAILABILITY

The supplementary information contains a discussion, using toy problems, on why we chose to develop our ML models with ridge regression, rather than other learning algorithms. It displays a table listing the SuperCon samples that we have identified as having T_c 's measured under applied pressure, and a table ranking the relative importance of each of the feature descriptors used in our ML models. It also displays two tables ranking the one-hundred highest T_c materials in the Materials Project identified by our implicit/ambient pressure models, as well as two additional tables obtained after subsequent filtering for band gaps lower than 1 eV.

Refer to <https://zenodo.org/records/11255989> for: Python code to generate ML features and to implement our similarity-based ML models; chemical compositions, T_c 's, pressures, ML features, and implicit/ambient pressure T_c predictions for our SuperCon samples; and chemical compositions, identifiers, energies above convex hulls, band gaps, ML features, and implicit/ambient pressure T_c predictions for materials in the Materials Project.

VI. ACKNOWLEDGMENTS

We acknowledge the support of the Natural Sciences and Engineering Research Council of Canada (NSERC), [funding reference number RGPIN-2023-04853]. Cette recherche a été financée par le Conseil de recherches

en sciences naturelles et en génie du Canada (CRSNG), [numéro de référence RGPIN-2023-04853]. This research was undertaken thanks in part to funding provided to the University of Toronto's Acceleration Consortium from the Canada First Research Excellence Fund, grant number: CFREF-2022-00042. O.A.v.L. has received support as the Ed Clark Chair of Advanced Materials and as a Canada CIFAR AI Chair.

VII. AUTHOR CONTRIBUTIONS

S.L. and O.A.v.L. conceived the idea. S.L. developed and implemented the methodology and performed all experiments, with guidance from O.A.v.L. All authors discussed the results, and made comments and edits to the manuscript written by S.L. and O.A.v.L.

-
- [1] A. Jain, S. P. Ong, G. Hautier, W. Chen, W. D. Richards, S. Dacek, S. Cholia, D. Gunter, D. Skinner, G. Ceder, *et al.*, *APL materials* **1** (2013).
- [2] "MDR SuperCon Datasheet // MDR — mdr.nims.go.jp," <https://mdr.nims.go.jp/collections/5712mb227> (2024), [Accessed 17-01-2024].
- [3] A. W. Sleight, *Accounts of chemical research* **28**, 103 (1995).
- [4] M. Tinkham, *Reviews of Modern Physics* **46**, 587 (1974).
- [5] L. Berger and B. Roberts, *CRC Handbook of Chemistry and Physics*, 12 (1997).
- [6] P. Mukherjee and V. Rao, *Physica C: Superconductivity and its applications* **563**, 67 (2019).
- [7] J. R. Hull, *Reports on Progress in Physics* **66**, 1865 (2003).
- [8] R. M. Scanlan, A. P. Malozemoff, and D. C. Larbalestier, *Proceedings of the IEEE* **92**, 1639 (2004).
- [9] W. V. Hassenzahl, D. W. Hazelton, B. K. Johnson, P. Komarek, M. Noe, and C. T. Reis, *Proceedings of the IEEE* **92**, 1655 (2004).
- [10] P. Bruzzone, *Physica C: Superconductivity and its applications* **470**, 1734 (2010).
- [11] P. Bruzzone, *Superconductor Science and Technology* **28**, 024001 (2014).
- [12] D. Whyte, J. Minervini, B. LaBombard, E. Marmor, L. Bromberg, and M. Greenwald, *Journal of Fusion Energy* **35**, 41 (2016).
- [13] A. Malozemoff, *Physica C: Superconductivity* **153**, 1049 (1988).
- [14] A. I. Braginski, *Journal of superconductivity and novel magnetism* **32**, 23 (2019).
- [15] K. K. Berggren, *Proceedings of the IEEE* **92**, 1630 (2004).
- [16] S. Bravyi, O. Dial, J. M. Gambetta, D. Gil, and Z. Nazario, *Journal of Applied Physics* **132** (2022).
- [17] J. R. Alonso and T. A. Antaya, *Reviews of Accelerator Science and Technology* **5**, 227 (2012).
- [18] S. Ali and M. Zulqarnain, *Superconductors: Materials and Applications* **132**, 211 (2022).
- [19] F. Werfel, U. Floegel-Delor, R. Rothfeld, T. Riedel, B. Goebel, D. Wippich, and P. Schirrmeyer, *Superconductor Science and Technology* **25**, 014007 (2011).
- [20] F. Werfel, U. Floegel-Delor, R. Rothfeld, T. Riedel, D. Wippich, B. Goebel, and P. Schirrmeyer, *Physics Procedia* **36**, 948 (2012).
- [21] L. Schultz, O. de Haas, P. Verges, C. Beyer, S. Rohlig, H. Olsen, L. Kuhn, D. Berger, U. Noteboom, and U. Funk, *IEEE Transactions on Applied Superconductivity* **15**, 2301 (2005).
- [22] M. Dam, R. Battiston, W. J. Burger, R. Carpentiero, E. Chesta, R. Iuppa, G. de Rijk, and L. Rossi, *Superconductor Science and Technology* **33**, 044012 (2020).
- [23] L. Rossi and L. Bottura, *Reviews of accelerator science and technology* **5**, 51 (2012).
- [24] A. Tollestrup and E. Todesco, *Reviews of Accelerator Science and Technology* **1**, 185 (2008).
- [25] L. Bottura, S. A. Gourlay, A. Yamamoto, and A. V. Zlobin, *IEEE Transactions on Nuclear Science* **63**, 751 (2015).
- [26] J. Bardeen, L. N. Cooper, and J. R. Schrieffer, *Physical review* **108**, 1175 (1957).
- [27] J. Bardeen, L. N. Cooper, and J. R. Schrieffer, *Physical Review* **106**, 162 (1957).
- [28] A. Mann, *Nature* **475**, 280 (2011).
- [29] D. Dew-Hughes, *Low temperature physics* **27**, 713 (2001).
- [30] A. Walsh, *Nature chemistry* **7**, 274 (2015).
- [31] A. Drozdov, P. Kong, V. Minkov, S. Besedin, M. Kuzovnikov, S. Mozaffari, L. Balicas, F. Balakirev, D. Graf, V. Prakapenka, *et al.*, *Nature* **569**, 528 (2019).
- [32] V. Stanev, C. Oses, A. G. Kusne, E. Rodriguez, J. Paglione, S. Curtarolo, and I. Takeuchi, *npj Computational Materials* **4**, 29 (2018).
- [33] T. Sommer, R. Willa, J. Schmalian, and P. Friederich, *Scientific Data* **10**, 816 (2023).
- [34] H. Tran and T. N. Vu, *Physical Review Materials* **7**, 054805 (2023).
- [35] L. Novakovic, A. Salamat, and K. V. Lawler, *arXiv preprint arXiv:2301.10474* (2023).
- [36] T. Konno, H. Kurokawa, F. Nabeshima, Y. Sakishita, R. Ogawa, I. Hosako, and A. Maeda, *Physical Review B* **103**, 014509 (2021).
- [37] P. Moscato, M. N. Haque, K. Huang, J. Sloan, and J. Corrales de Oliveira, *Algorithms* **16**, 382 (2023).
- [38] B. Meredig, E. Antoni, C. Church, M. Hutchinson, J. Ling, S. Paradiso, B. Blaiszik, I. Foster, B. Gibbons, J. Hattrick-Simpers, *et al.*, *Molecular Systems Design & Engineering* **3**, 819 (2018).
- [39] C. Pereti, K. Bernot, T. Guizouarn, F. Laufek, A. Vymazalová, L. Bindi, R. Sessoli, and D. Fanelli, *Npj Computational Materials* **9**, 71 (2023).
- [40] P. W. Koh, S. Sagawa, H. Marklund, S. M. Xie, M. Zhang, A. Balsubramani, W. Hu, M. Yasunaga, R. L. Phillips, I. Gao, *et al.*, in *International Conference on Machine Learning* (PMLR, 2021) pp. 5637–5664.
- [41] Y. Wald, A. Feder, D. Greenfeld, and U. Shalit, *Advances in neural information processing systems* **34**, 2215 (2021).

- [42] I. Gulrajani and D. Lopez-Paz, arXiv preprint arXiv:2007.01434 (2020).
- [43] M. Akrout, A. Feriani, F. Bellili, A. Mezghani, and E. Hossain, IEEE Communications Surveys & Tutorials (2023).
- [44] L. Foppiano, P. B. Castro, P. Ortiz Suarez, K. Terashima, Y. Takano, and M. Ishii, Science and Technology of Advanced Materials: Methods **3**, 2153633 (2023).
- [45] P. Dai, B. Chakoumakos, G. Sun, K. Wong, Y. Xin, and D. Lu, Physica C: Superconductivity **243**, 201 (1995).
- [46] D. Djurek, Z. Medunić, A. Tonejc, and M. Paljević, Physica C: Superconductivity **351**, 78 (2001).
- [47] E. Snider, N. Dasenbrock-Gammon, R. McBride, M. Debessai, H. Vindana, K. Vencatasamy, K. V. Lawler, A. Salamat, and R. P. Dias, Nature **588**, E18 (2020).
- [48] P. Ayyub, P. Guptasarma, A. Rajarajan, L. Gupta, R. Vijayaraghavan, and M. Multani, Journal of Physics C: Solid State Physics **20**, L673 (1987).
- [49] D. Henshaw, D. Hurst, and N. Pope, Physical Review **92**, 1229 (1953).
- [50] A. Drozdov, M. Eremets, I. Troyan, V. Ksenofontov, and S. I. Shylin, Nature **525**, 73 (2015).
- [51] K. Shimizu, M. Einaga, M. Sakata, A. Masuda, H. Nakao, M. Eremets, A. Drozdov, I. Troyan, N. Hirao, S. Kawaguchi, *et al.*, Physica C: Superconductivity and its applications **552**, 27 (2018).
- [52] H. Shao, C. Lam, P. Fung, X. Wu, J. Du, G. Shen, J. Chow, S. Ho, K. Hung, and X. Yao, Physica C: Superconductivity **246**, 207 (1995).
- [53] D. Lemm, G. F. von Rudorff, and O. A. von Lilienfeld, Machine Learning: Science and Technology **4**, 045043 (2023).
- [54] W. N. van Wieringen, arXiv preprint arXiv:1509.09169 (2015).
- [55] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and E. Duchesnay, Journal of Machine Learning Research **12**, 2825 (2011).
- [56] “sklearn.linear_model.Ridge — scikit-learn.org,” https://scikit-learn.org/stable/modules/generated/sklearn.linear_model.Ridge.html (2024).
- [57] “sklearn.linear_model.RidgeCV — scikit-learn.org,” https://scikit-learn.org/stable/modules/generated/sklearn.linear_model.RidgeCV.html (2024).
- [58] B. Meredig, A. Agrawal, S. Kirklin, J. E. Saal, J. W. Doak, A. Thompson, K. Zhang, A. Choudhary, and C. Wolverton, Physical Review B **89**, 094104 (2014).
- [59] L. M. Ghiringhelli, J. Vybiral, S. V. Levchenko, C. Draxl, and M. Scheffler, Physical review letters **114**, 105503 (2015).
- [60] D. Jha, L. Ward, A. Paul, W.-k. Liao, A. Choudhary, C. Wolverton, and A. Agrawal, Scientific reports **8**, 17593 (2018).
- [61] R. E. Goodall and A. A. Lee, Nature communications **11**, 6280 (2020).
- [62] A. Y.-T. Wang, S. K. Kauwe, R. J. Murdock, and T. D. Sparks, Npj Computational Materials **7**, 77 (2021).
- [63] J. Damewood, J. Karaguesian, J. R. Lunger, A. R. Tan, M. Xie, J. Peng, and R. Gómez-Bombarelli, Annual Review of Materials Research **53**, 399 (2023).
- [64] L. Ward, A. Dunn, A. Faghaninia, N. E. Zimmermann, S. Bajaj, Q. Wang, J. Montoya, J. Chen, K. Bystrom, M. Dylla, *et al.*, Computational Materials Science **152**, 60 (2018).
- [65] C. Cortes, L. D. Jackel, S. Solla, V. Vapnik, and J. Denker, Advances in neural information processing systems **6** (1993).
- [66] P. Pernot, B. Huang, and A. Savin, Machine Learning: Science and Technology **1**, 035011 (2020).
- [67] E. Maxwell, Physical Review **78**, 477 (1950).
- [68] C. Reynolds, B. Serin, W. Wright, and L. Nesbitt, Physical Review **78**, 487 (1950).
- [69] V. Z. Kresin, A. Bill, S. A. Wolf, and Y. N. Ovchinnikov, Physical Review B **56**, 107 (1997).
- [70] G.-m. Zhao, H. Keller, and K. Conder, Journal of Physics: Condensed Matter **13**, R569 (2001).
- [71] A. D'yachenko, V. Y. Tarenkov, R. Szymczak, A. Abal'oshev, I. Abal'osheva, S. Lewandowski, and L. Leonyuk, Physical Review B **61**, 1500 (2000).
- [72] F. Flacke and H. Jacobs, Journal of alloys and compounds **227**, 109 (1995).
- [73] M. Project, “mp-505233,” [https://next-gen.materialsproject.org/materials/mp-505233?](https://next-gen.materialsproject.org/materials/mp-505233?material_ids=mp-505233) (), [Accessed April 2024].
- [74] V. Khodel and V. Shaginyan, Jetp Lett **51**, 553 (1990).
- [75] T. T. Heikkilä and G. E. Volovik, Basic Physics of Functionalized Graphite , 123 (2016).
- [76] R. W. Godby, M. Schlüter, and L. Sham, Physical Review B **37**, 10159 (1988).
- [77] M. Project, “Electronic Structure — Materials Project Documentation — docs.materialsproject.org,” <https://docs.materialsproject.org/methodology/materials-methodology/electronic-structure> (), [Accessed April 2024].
- [78] V. Vovk, in *Empirical Inference: Festschrift in Honor of Vladimir N. Vapnik* (Springer, 2013) pp. 105–116.
- [79] T. Chen and C. Guestrin, in *Proceedings of the 22nd ACM SIGKDD international conference on knowledge discovery and data mining* (2016) pp. 785–794.
- [80] E. A. Nadaraya, Theory of Probability & Its Applications **9**, 141 (1964).
- [81] G. S. Watson, Sankhyā: The Indian Journal of Statistics, Series A , 359 (1964).
- [82] H. J. Bierens, (1988).
- [83] A. Malistov and A. Trushin, in *2019 18th IEEE International Conference On Machine Learning And Applications (ICMLA)* (IEEE, 2019) pp. 783–789.
- [84] T. K. Ho, in *Proceedings of 3rd international conference on document analysis and recognition*, Vol. 1 (IEEE, 1995) pp. 278–282.
- [85] T. K. Ho, IEEE transactions on pattern analysis and machine intelligence **20**, 832 (1998).

VIII. SUPPLEMENTARY INFORMATION

A. Toy Problem

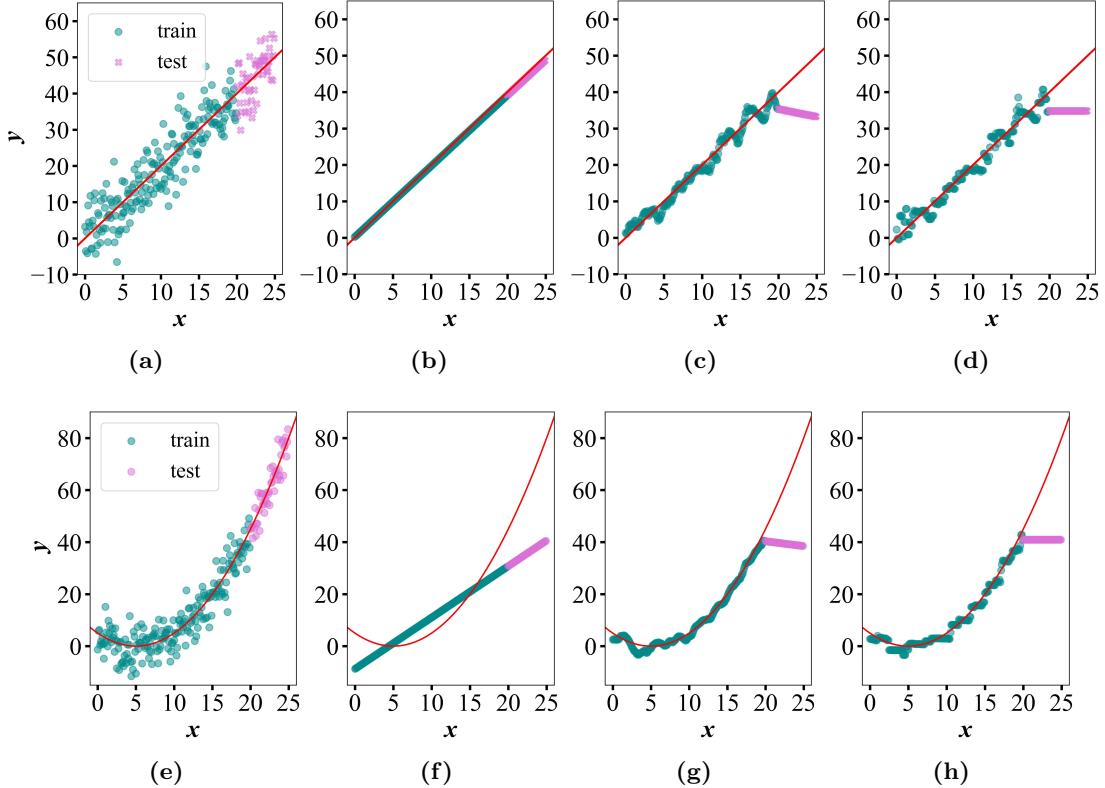


FIG. 7. A simple toy problem of (a) $y = 2x$ with the addition of Gaussian noise, where the label values of the training set are less than those of the test set. Shown are the predictions of three models on the training and test sets, developed with the following three popular machine learning algorithms: (b) ridge regression (c) kernel ridge regression and (d) eXtreme Gradient Boosting regression. Similarly, the exercise is repeated for (e) $y = 0.2(x - 5)^2$, with model predictions obtained using (f) ridge regression (g) kernel ridge regression (h) and eXtreme Gradient Boosting regression.

The selection of ridge regression over other learning algorithms in our ML models is motivated by its ability to potentially make accurate predictions for samples with labels found beyond the distribution of the training set. Consider the toy example of $y = 2x$ with the addition of Gaussian noise in Figure 7a. Here, the training set contains feature and label values that are both less than those found in the test set. Thus, with this OOD transfer learning problem, the performances of ML models trained with different learning algorithms (hyperparameters selected as those returning the lowest mean absolute error, MAE, upon five-fold cross-validation on the training set) are evaluated by their abilities to effectively extrapolate beyond their training set.

Ridge regression (Figure 7b) learns the linearity of the problem since for greater values of x , it correctly predicts correspondingly greater values of y in both the training and test sets, although its accuracy could be improved. Models developed using two other popular and powerful ML algorithms of kernel ridge regression⁷⁸ with the Laplacian kernel (Figure 7c) and eXtreme Gradient Boosting⁷⁹ (XGBoost) regression with the tree-based `gbtree` booster (Figure 7d) exhibit, relative to the ridge regression model, improved performances on the training set. However, regarding the test set, both fail to make even a single prediction of y that is greater than the maximum value found in the training set. This is, perhaps, an unsurprising result. After all, the objective of a kernel-based regression algorithm is to find a non-parametric mapping between the domain of feature vectors, \mathbf{X} , and their labels, \mathbf{Y} , where the estimated function outputs, $\hat{\mathbf{Y}}$, are obtained as kernel-weighted local averages^{78,80–82}. Thus, this fundamentally restricts the value that a given $\hat{\mathbf{Y}}$ can take to within the range of values found in $\mathbf{Y}_{\text{train}}$. The same conclusion results from a tree-based regression algorithm, which returns the weighted-average of the predictions obtained from an ensemble of decision trees^{83–85}.

In contrast, ridge regression weights a sample's feature values, which thus does not limit the model's output to within the distribution of \mathbf{Y}_{train} . Therefore, ridge regression appears to be suitable for the objective of predicting label values beyond those found in the training set. Of course, before applying it to the prediction of out-of-distribution test samples, care must be taken to ensure that its estimates are robust and accurate since it requires the assumption that there exists some linear relationship between the features and labels space. If this condition is not fulfilled, then it can be seen from a quadratic toy problem of $y = 0.2(x - 5)^2$ ([Figure 7e](#)) that ridge regression exhibits poor predictive power in both the train and test regimes due to the non-linearity of the problem. The fundamental limits on the labels space imposed by the different learning algorithms are observed again, since ridge regression can predict y_{test} values greater than those found in training, albeit poorly in this case; kernel ridge regression ([Figure 7g](#)) and XGBoost regression ([Figure 7h](#)) exhibit good performances on the training set but they fail once more to predict a single value of y_{test} beyond \mathbf{Y}_{train} .

B. SuperCon Pressure Samples

TABLE III. Compositions of materials in SuperCon that we have identified as having T_c 's measured under applied pressure.

Composition	T_c (K)
IrU	0
SnTe	0
KC ₈	0
Zn ₂ Zr	0
Ge ₂ U	0
LaC ₂	1
Te ₂ U	1
PdTe ₂	1
TiSe ₂	1
As ₂ Cd ₃	3
OSn	3
S ₂ Ta	3
BaBi ₃	6
B ₁₂ Zr	6
C ₆ Yb	6
Au ₂ Pb	6
Te ₂ W	7
Sb ₂ Te ₃	7
NaAlSi	7
Mo ₃ Sb ₇	7
MoTe ₂	8
B ₂ Nb	9
C ₆ Ca	11
H ₄ Si	17
NNb	17
C ₃ Y ₂	18
LiFeAs	18
Nb ₃ Sn	18
GeNb ₃	22
NaFeAs	26
B ₂ Mg	40
FeSe	80
Hg _{0.75} Ba _{2.07} Ca _{2.07} Cu _{3.11} O _{8.208}	135
Hg _{0.66} Pb _{0.34} Ba ₂ Ca _{1.98} Cu _{2.9} O _{8.4}	143
H ₃ S	147
H ₂ S	203
LaH ₁₀	250

C. Features Rankings

TABLE IV: Ranking of the relative average contribution of each ML feature to the T_c predictions in the SuperCon implicit/ambient pressure models, from greatest to least.

Ranking	Implicit pressure	Ambient pressure
1	Atomic number, mean	Row number, mean
2	Covalent radius, mean	Atomic number, mean
3	Atomic weight, mean	Atomic weight, mean
4	Atomic number, avg. dev.	Atomic number, avg. dev.
5	Atomic weight, avg. dev.	Atomic weight, avg. dev.
6	Mendeleev number, mean	Atomic number, max.
7	Atomic number, mode	Atomic weight, max.
8	Atomic weight, mode	2-norm
9	Column number, mean	Electronegativity, mean
10	Atomic number, max.	Covalent radius, mean
11	Atomic weight, max.	Column number, mean
12	Row number, mean	Column number, mode
13	Elemental solid volume, mean	Mendeleev number, mean
14	Covalent radius, max.	Covalent radius, max.
15	Covalent radius, mode	7-norm
16	Column number, max.	5-norm
17	Column number, mode	3-norm
18	Mendeleev number, max.	Atomic number, mode
19	Atomic weight, range	10-norm
20	Electronegativity, mean	Row number, mode
21	Atomic number, range	Mendeleev number, mode
22	Electronegativity, mode	Mendeleev number, max.
23	Mendeleev number, mode	Column number, avg. dev.
24	Covalent radius, min.	Number of valence electrons, mean
25	Row number, max.	Atomic weight, mode
26	Row number, mode	Atomic weight, range
27	Number of valence electrons, mean	Column number, max.
28	Column number, range	Atomic number, range
29	Number of unfilled valence orbitals, mean	Number of unfilled valence orbitals, mean
30	Space group number, mean	Atomic number, min.
31	Melting temperature, mean	Number of <i>s</i> valence electrons, mean
32	Covalent radius, range	Number of filled <i>s</i> valence orbitals, mean
33	Column number, avg. dev.	Covalent radius, range
34	Elemental solid volume, mode	Electronegativity, max.
35	Atomic number, min.	Space group number, max.
36	Space group number, max.	Electronegativity, mode
37	Mendeleev number, avg. dev.	Atomic weight, min.
38	0-norm	Elemental solid volume, mean
39	Covalent radius, avg. dev.	Column number, range
40	Number of valence electrons, mode	Covalent radius, mode
41	Number of <i>s</i> valence electrons, mean	Number of filled <i>p</i> valence orbitals, max.
42	Number of filled <i>s</i> valence orbitals, mean	Number of valence electrons, mode
43	Atomic weight, min.	Mendeleev number, avg. dev.
44	Melting temperature, avg. dev.	Number of filled <i>s</i> valence orbitals, max.
45	Elemental solid volume, max.	Elemental solid volume, max.
46	Electronegativity, max.	Number of filled <i>s</i> valence orbitals, mode
47	Melting temperature, max.	Electronegativity, avg. dev.
48	2-norm	Space group number, mean
49	Mendeleev number, range	Covalent radius, min.
50	Row number, min.	0-norm
51	Number of unfilled valence orbitals, mode	Row number, min.
52	Number of filled <i>p</i> valence orbitals, avg. dev.	Melting temperature, max.
53	Space group number, avg. dev.	Frac. <i>d</i> valence electrons
54	Melting temperature, range	Row number, avg. dev.

Continued on next page

TABLE IV: (Continued)

Ranking	Implicit pressure	Ambient pressure
55	Melting temperature, mode	Elemental solid volume, mode
56	Number of filled d valence orbitals, max.	Mendeleev number, range
57	Number of filled d valence orbitals, mean	Row number, max.
58	Number of d valence electrons, mean	Number of filled p valence orbitals, range
59	Row number, range	Number of filled p valence orbitals, avg. dev.
60	Elemental solid volume, avg. dev.	Covalent radius, avg. dev.
61	Number of filled p valence orbitals, max.	Electronegativity, min.
62	Number of filled s valence orbitals, mode	Number of unfilled valence orbitals, mode
63	3-norm	Frac. s valence electrons
64	Row number, avg. dev.	Row number, range
65	Electronegativity, min.	Number of unfilled valence orbitals, max.
66	Elemental solid volume, range	Elemental solid volume, range
67	Column number, min.	Number of unfilled d valence orbitals, avg. dev.
68	Number of valence electrons, max.	Number of filled d valence orbitals, avg. dev.
69	Number of filled d valence orbitals, avg. dev.	Number of valence electrons, max.
70	10-norm	Frac. p valence electrons
71	Number of valence electrons, avg. dev.	Elemental solid volume, min.
72	Number of filled d valence orbitals, range	Melting temperature, avg. dev.
73	Number of unfilled valence orbitals, max.	Space group number, range
74	Number of filled p valence orbitals, range	Melting temperature, mean
75	Space group number, range	Number of filled p valence orbitals, mean
76	7-norm	Number of p valence electrons, mean
77	Frac. d valence electrons	Electronegativity, range
78	Space group number, mode	Column number, min.
79	Number of unfilled p valence orbitals, avg. dev.	Number of unfilled p valence orbitals, mean
80	Number of unfilled p valence orbitals, mean	Number of d valence electrons, mean
81	Number of filled s valence orbitals, max.	Number of filled d valence orbitals, mean
82	Elemental solid volume, min.	Number of unfilled d valence orbitals, mean
83	5-norm	Number of unfilled p valence orbitals, max.
84	Electronegativity, avg. dev.	Space group number, mode
85	Mendeleev number, min.	Number of unfilled valence orbitals, avg. dev.
86	Number of unfilled d valence orbitals, mean	Mendeleev number, min.
87	Number of unfilled valence orbitals, avg. dev.	Melting temperature, mode
88	Number of unfilled d valence orbitals, avg. dev.	Melting temperature, range
89	Space group number, min.	Number of valence electrons, avg. dev.
90	Number of valence electrons, range	Number of valence electrons, range
91	Number of filled p valence orbitals, mean	Number of filled d valence orbitals, range
92	Number of p valence electrons, mean	Number of filled s valence orbitals, min.
93	Electronegativity, range	Number of filled d valence orbitals, max.
94	Number of filled s valence orbitals, min.	Number of unfilled p valence orbitals, range
95	Number of unfilled valence orbitals, range	Number of unfilled s valence orbitals, avg. dev.
96	Melting temperature, min.	Space group number, avg. dev.
97	Number of unfilled p valence orbitals, max.	Number of unfilled valence orbitals, range
98	Frac. s valence electrons	Number of unfilled d valence orbitals, max.
99	Number of unfilled p valence orbitals, range	Number of filled p valence orbitals, mode
100	Number of unfilled d valence orbitals, max.	MagpieData avg_dev GSvolume.pa
101	Number of filled p valence orbitals, mode	Number of unfilled s valence orbitals, range
102	Number of valence electrons, min.	Number of unfilled p valence orbitals, avg. dev.
103	Number of unfilled p valence orbitals, mode	Number of valence electrons, min.
104	Number of unfilled d valence orbitals, range	Number of unfilled p valence orbitals, mode
105	Transition metal fraction	Number of filled s valence orbitals, range
106	Frac. p valence electrons	Space group number, min.
107	Number of filled s valence orbitals, range	Number of filled s valence orbitals, avg. dev.
108	Number of unfilled s valence orbitals, range	Number of unfilled s valence orbitals, max.

Continued on next page

TABLE IV: (Continued)

Ranking	Implicit pressure	Ambient pressure
109	Number of unfilled <i>s</i> valence orbitals, avg. dev.	Melting temperature, min.
110	Number of filled <i>f</i> valence orbitals, max.	Number of filled <i>d</i> valence orbitals, mode
111	Number of unfilled <i>s</i> valence orbitals, max.	Frac. <i>f</i> valence electrons
112	Number of filled <i>s</i> valence orbitals, avg. dev.	Transition metal fraction
113	Number of filled <i>f</i> valence orbitals, avg. dev.	Number of unfilled <i>s</i> valence orbitals, mean
114	Number of filled <i>d</i> valence orbitals, mode	Number of unfilled <i>d</i> valence orbitals, range
115	Number of <i>f</i> valence electrons, mean	Elemental solid band gap, max.
116	Number of filled <i>f</i> valence orbitals, mean	Number of filled <i>f</i> valence orbitals, avg. dev.
117	Number of unfilled <i>s</i> valence orbitals, mean	Elemental solid band gap, range
118	Number of filled <i>f</i> valence orbitals, range	Number of filled <i>f</i> valence orbitals, max.
119	Frac. <i>f</i> valence electrons	Number of filled <i>f</i> valence orbitals, range
120	Number of unfilled valence orbitals, min.	Number of unfilled <i>d</i> valence orbitals, mode
121	Elemental solid band gap, mean	Elemental solid band gap, mean
122	Elemental solid band gap, avg. dev.	Elemental solid band gap, avg. dev.
123	Elemental solid band gap, max.	Number of unfilled valence orbitals, min.
124	Elemental solid band gap, range	Number of <i>f</i> valence electrons, mean
125	Number of unfilled <i>d</i> valence orbitals, mode	Number of filled <i>f</i> valence orbitals, mean
126	Elemental solid magnetic moment, avg. dev.	Elemental solid magnetic moment, avg. dev.
127	Number of unfilled <i>f</i> valence orbitals, avg. dev.	Elemental solid magnetic moment, mean
128	Number of unfilled <i>f</i> valence orbitals, mean	Number of unfilled <i>f</i> valence orbitals, avg. dev.
129	Elemental solid magnetic moment, mean	Number of filled <i>d</i> valence orbitals, min.
130	Number of filled <i>d</i> valence orbitals, min.	Number of unfilled <i>f</i> valence orbitals, mean
131	Number of unfilled <i>f</i> valence orbitals, range	Elemental solid band gap, mode
132	Number of unfilled <i>f</i> valence orbitals, max.	Number of unfilled <i>s</i> valence orbitals, mode
133	Elemental solid band gap, mode	Elemental solid magnetic moment, max.
134	Elemental solid magnetic moment, range	Elemental solid magnetic moment, range
135	Elemental solid magnetic moment, max.	Number of unfilled <i>f</i> valence orbitals, max.
136	Number of unfilled <i>s</i> valence orbitals, mode	Number of unfilled <i>f</i> valence orbitals, range
137	Number of unfilled <i>d</i> valence orbitals, min.	Number of unfilled <i>d</i> valence orbitals, min.
138	Number of filled <i>f</i> valence orbitals, mode	Number of filled <i>f</i> valence orbitals, mode
139	Number of unfilled <i>p</i> valence orbitals, min.	Number of filled <i>p</i> valence orbitals, min.
140	Number of filled <i>p</i> valence orbitals, min.	Number of unfilled <i>p</i> valence orbitals, min.
141	Elemental solid magnetic moment, mode	Elemental solid magnetic moment, mode
142	Number of filled <i>f</i> valence orbitals, min.	Number of filled <i>f</i> valence orbitals, min.
143	Number of unfilled <i>s</i> valence orbitals, min.	Number of unfilled <i>s</i> valence orbitals, min.
144	Number of unfilled <i>f</i> valence orbitals, mode	Number of unfilled <i>f</i> valence orbitals, mode
145	Elemental solid band gap, min.	Elemental solid band gap, min.
146	Elemental solid magnetic moment, min.	Elemental solid magnetic moment, min.
147	Number of unfilled <i>f</i> valence orbitals, min.	Number of unfilled <i>f</i> valence orbitals, min.

D. Materials Project, Implicit Pressure

TABLE V: Top 100 materials in the Materials Project with highest T_c predictions under unknown (potentially very high) pressure, dot product spreads of less than 51 K, and energies above their convex hulls of less than 0.030 eV/atom.

	ID	Composition	Predicted T_c (K)	Dot product spread (K)	Hull energy (eV/atom)
1	mp-505233	$\text{Cs}_2\text{Sn}(\text{H}_2\text{N})_6$	324	48	0.017
2	mp-1184037	Csh_5N_2	315	47	0.011
3	mp-1198227	Csh_5N_2	315	47	0.000
4	mp-643359	$\text{Rb}_2\text{Sn}(\text{H}_2\text{N})_6$	305	45	0.028
5	mp-643371	$\text{K}_2\text{Sn}(\text{H}_2\text{N})_6$	301	43	0.000
6	mp-1193046	$\text{CsMgAs}(\text{H}_6\text{O}_5)_2$	297	35	0.000
7	mp-643158	PH_8IN_4	297	46	0.000
8	mp-974267	KH_8N_3	288	35	0.000
9	mp-721084	H_7IN_2	287	47	0.000
10	mp-1202629	$\text{CdH}_{20}\text{N}_6\text{OF}_2$	286	45	0.000
11	mp-1204788	$\text{InH}_{12}(\text{NF}_2)_3$	279	44	0.000
12	mp-738629	$\text{CsGaH}_{24}(\text{SO}_{10})_2$	277	46	0.000
13	mp-1198073	CsH_4NF_2	275	48	0.004
14	mp-767240	$\text{CsPH}_4(\text{NO})_2$	274	47	0.000
15	mp-24113	$\text{SnH}_8(\text{NF}_3)_2$	273	45	0.000
16	mp-1196740	$\text{Cs}_3\text{H}_{12}\text{N}_4\text{F}_3$	271	48	0.000
17	mp-863000	$\text{SnH}_6(\text{NF}_2)_2$	270	45	0.000
18	mp-758953	$\text{P}_2\text{H}_{12}\text{BrN}_7$	269	40	0.008
19	mp-642740	$\text{CsLi}(\text{H}_2\text{N})_2$	269	45	0.000
20	mp-1194837	$\text{Ba}_3\text{As}_2\text{H}_{34}\text{O}_{25}$	268	34	0.024
21	mp-1197059	$\text{LiSn}(\text{H}_2\text{N})_3$	266	44	0.000
22	mp-1198991	$\text{Rb}_3\text{H}_{12}\text{N}_5$	265	40	0.000
23	mp-677248	$\text{Ca}_2\text{AlH}_{10}\text{IO}_8$	265	37	0.000
24	mp-643902	$\text{SnH}_4(\text{NF})_2$	262	46	0.000
25	mp-733457	$\text{CsNa}_2(\text{H}_5\text{O}_3)_3$	261	44	0.003
26	mp-733932	$\text{InH}_5(\text{NF})_2$	260	45	0.000
27	mp-706544	$\text{NaGa}(\text{H}_2\text{N})_4$	258	38	0.006
28	mp-28204	CsH_7O_4	257	46	0.000
29	mp-1179724	$\text{RbLi}_2(\text{H}_2\text{N})_3$	256	37	0.006
30	mp-722455	$\text{RbLi}_2(\text{H}_2\text{N})_3$	256	37	0.006
31	mp-643394	$\text{NaInH}_8(\text{NF}_3)_2$	255	44	0.028
32	mp-866716	$\text{RbLiH}_{12}\text{Se}_3\text{N}_4$	255	42	0.000
33	mp-849391	$\text{InH}_9(\text{NCl})_3$	254	45	0.000
34	mp-28892	PH_4N_3	253	36	0.000
35	mp-7714	Li_4UO_5	251	48	0.000
36	mp-559463	$\text{RbMgAs}(\text{H}_6\text{O}_5)_2$	250	31	0.000
37	mp-34381	H_4IN	248	50	0.006
38	mp-643062	H_4IN	248	50	0.000
39	mp-505786	$\text{CeH}_{14}\text{Cl}_3\text{O}_7$	246	37	0.011
40	mp-1195832	$\text{SiH}_{12}(\text{NF})_4$	246	41	0.012
41	mp-1204167	$\text{CeH}_{14}\text{Cl}_3\text{O}_7$	246	37	0.014
42	mp-1198095	$\text{CsAlH}_{24}(\text{SO}_{10})_2$	245	35	0.000
43	mp-556009	$\text{MgTlAs}(\text{H}_6\text{O}_5)_2$	244	41	0.000
44	mp-1192659	$\text{LiGa}(\text{H}_2\text{N})_4$	243	37	0.002
45	mp-626263	$\text{Ba}(\text{H}_8\text{O}_5)_2$	243	32	0.001
46	mp-626268	$\text{Ba}(\text{H}_8\text{O}_5)_2$	243	32	0.000
47	mp-626264	$\text{Ba}(\text{H}_8\text{O}_5)_2$	243	32	0.026
48	mp-626297	$\text{Ba}(\text{H}_8\text{O}_5)_2$	243	32	0.028
49	mp-1353537	$\text{Ba}(\text{H}_8\text{O}_5)_2$	243	32	0.001
50	mp-1194674	$\text{LiGa}(\text{H}_2\text{N})_4$	243	37	0.000
51	mp-722774	$\text{Ba}_3\text{As}_2\text{H}_{14}\text{S}_8\text{O}_7$	243	37	0.013

Continued on next page

TABLE V: (Continued)

	ID	Composition	Predicted T_c (K)	Dot product spread (K)	Hull energy (eV/atom)
52	mp-1182486	Ba(H ₈ O ₅) ₂	243	32	0.024
53	mp-1202018	CsH ₄ NO	242	48	0.000
54	mp-1201190	GaH ₉ (NF) ₃	241	40	0.000
55	mp-510073	RbLi(H ₂ N) ₂	241	38	0.000
56	mp-707956	KLi ₇ (H ₂ N) ₈	240	32	0.022
57	mp-773582	LiTe(H ₃ N) ₄	240	37	0.000
58	mp-1202913	GaH ₁₅ N ₅ Cl ₃	240	40	0.000
59	mp-1202249	Na ₃ Np(H ₇ O ₆) ₂	240	44	0.016
60	mp-1195064	Sn ₂ H ₂₂ S ₆ N ₄ O ₃	239	45	0.000
61	mp-703559	KSrAs(H ₄ O ₃) ₄	238	31	0.000
62	mp-1196329	K ₂ AsH ₁₃ O ₁₀	237	30	0.000
63	mp-1202324	P ₂ H ₁₀ SN ₆	237	36	0.004
64	mp-761178	CsMgP(H ₆ O ₅) ₂	237	35	0.000
65	mp-541005	CsH ₅ O ₃	236	47	0.000
66	mp-1194535	CaH ₁₆ (IO ₄) ₂	236	37	0.010
67	mp-1195969	CaH ₁₄ I ₂ O ₇	236	37	0.012
68	mp-8609	Li ₆ UO ₆	236	50	0.000
69	mp-28587	BaH ₈ O ₅	236	30	0.000
70	mp-1196878	P ₂ H ₁₀ N ₆ O	236	36	0.015
71	mp-1212119	KCaAs(H ₄ O ₃) ₄	236	32	0.005
72	mp-758356	P ₂ H ₁₂ N ₇ Cl	236	37	0.009
73	mp-24523	CsAlH ₂₄ (SeO ₁₀) ₂	235	35	0.000
74	mp-634446	CsAl(H ₂ N) ₄	235	31	0.000
75	mp-1195610	Ca ₂ H ₂₆ I ₄ O ₁₃	234	37	0.003
76	mp-707454	Li(H ₃ N) ₄	234	34	0.028
77	mp-760046	LiPH ₂₁ S ₃ N ₇	233	36	0.000
78	mp-1224176	KMg ₂ As ₂ H ₃₁ O ₂₃	233	30	0.017
79	mp-781992	PH ₈ N ₄ Cl	232	37	0.005
80	mp-1194478	CsP ₂ H ₃ N	232	48	0.023
81	mp-758253	CaH ₁₂ (IO ₆) ₂	232	39	0.024
82	mp-1226730	CdH ₆ (NCl) ₂	230	47	0.000
83	mp-760697	GaH ₇ (NF ₂) ₂	230	40	0.019
84	mp-722502	Li ₃ P ₁₁ (H ₃ N) ₁₇	230	35	0.007
85	mp-1200428	KMgAs(H ₆ O ₅) ₂	230	30	0.013
86	mp-865095	NaGaH ₈ (NF ₃) ₂	229	39	0.008
87	mp-23675	H ₄ BrN	229	42	0.000
88	mp-36248	H ₄ BrN	229	42	0.001
89	mp-1224894	GaH ₆ N ₂ F ₃	229	40	0.000
90	mp-695316	Ca ₂ AlH ₁₀ BrO ₈	228	31	0.000
91	mp-1195558	Zn(H ₁₅ N ₈) ₂	228	28	0.000
92	mp-1200443	CaSnP ₆ (HO) ₁₂	228	39	0.013
93	mp-721171	CsMgH ₁₂ (ClO ₂) ₃	228	35	0.000
94	mp-1227507	Ca ₂ AlH ₁₀ BrO ₈	228	31	0.001
95	mp-696275	SnH ₈ (NCl ₃) ₂	227	47	0.007
96	mp-1195052	CaNp(HO) ₉	227	42	0.000
97	mp-23763	SnH ₈ (NCl ₃) ₂	227	47	0.000
98	mp-707324	PH ₆ SN ₃	226	36	0.000
99	mp-706979	PH ₆ N ₃ O	226	36	0.007
100	mp-759242	InH ₁₀ N ₂ Cl ₅ O	226	46	0.011

E. Materials Project, Implicit Pressure, Small Band Gaps

TABLE VI: Top 100 materials in the Materials Project with the highest T_c predictions under unknown (potentially very high) pressure, dot product spreads of less than 51 K, energies above their convex hulls of less than 0.030 eV/atom, and band gaps of less than 1.000 eV.

	ID	Composition	Predicted T_c (K)	Dot product spread (K)	Hull energy (eV/atom)	Band gap (eV)
1	mp-866716	RbLiH ₁₂ Se ₃ N ₄	255	42	0.000	0.801
2	mp-505786	CeH ₁₄ Cl ₃ O ₇	246	37	0.011	0.023
3	mp-1204167	CeH ₁₄ Cl ₃ O ₇	246	37	0.014	0.328
4	mp-707454	Li(H ₃ N) ₄	234	34	0.028	0.000
5	mp-1198097	Sr ₂ SnH ₁₈ (SeO ₃) ₄	212	34	0.000	0.819
6	mp-721650	CaH ₁₄ I ₁₀ O ₇	198	45	0.022	0.678
7	mp-570931	SrLi ₃ MnN ₃	192	47	0.000	0.000
8	mp-1202248	ReH ₃₀ Ru ₂ (NCl) ₁₀	180	38	0.029	0.031
9	mp-643738	Eu(MgH ₃) ₂	176	34	0.000	0.000
10	mp-1181729	Eu ₂ Mg ₃ H ₁₀	174	35	0.020	0.442
11	mp-541365	LiEuH ₃	172	38	0.000	0.000
12	mp-697962	Eu ₆ Mg ₇ H ₂₆	169	35	0.014	0.427
13	mp-1245450	Ca ₈ (MnN ₃) ₃	168	46	0.020	0.000
14	mp-697677	CeMg ₂ H ₇	167	32	0.000	0.000
15	mp-643756	EuMgH ₄	164	36	0.000	0.171
16	mp-696588	PH ₃	158	36	0.029	0.000
17	mp-973064	LaH ₃	154	45	0.003	0.471
18	mp-1018144	LaH ₃	154	45	0.000	0.000
19	mp-569112	Li ₃ CaMnN ₃	152	47	0.000	0.073
20	mp-754586	Li ₈ BiO ₆	148	48	0.029	0.000
21	mp-1177487	Li ₄₇ (CoO ₄) ₈	144	49	0.023	0.828
22	mp-543090	Li ₁₇ (CoO ₄) ₃	140	48	0.019	0.000
23	mp-543091	Li ₁₆ (CoO ₄) ₃	140	45	0.025	0.000
24	mp-1172980	Li ₁₇ (CoO ₄) ₃	140	48	0.015	0.345
25	mp-1185524	Li ₉₅ Mn ₁₆ O ₆₄	138	31	0.013	0.000
26	mp-677406	CuH ₁₂ N ₂ (Cl ₂ O) ₂	138	40	0.024	0.694
27	mp-604996	CuH ₁₂ N ₂ (Cl ₂ O) ₂	138	40	0.014	0.655
28	mp-721415	CuH ₁₂ N ₂ (Cl ₂ O) ₂	138	40	0.018	0.533
29	mp-989535	Cs ₂ AlInH ₆	137	36	0.000	0.611
30	mp-769483	Li ₁₁ (CoO ₄) ₂	135	44	0.024	0.289
31	mp-1188721	IN ₄	135	50	0.000	0.000
32	mp-699498	U ₂ CuP ₂ (HO) ₂₄	132	50	0.000	0.918
33	mp-1202929	AgH ₈ C ₂ S ₂ N ₄ Cl	131	35	0.024	0.000
34	mp-1102496	EuH ₂	130	39	0.000	0.000
35	mp-1229294	AlZnH ₁₀ O ₇ F ₅	129	27	0.022	0.990
36	mp-1183713	CeH ₃	126	35	0.003	0.000
37	mp-698125	NpH ₁₆ C ₄ N ₈ ClO ₆	126	37	0.010	0.429
38	mp-22950	CsNiCl ₃	125	45	0.000	0.805
39	mp-1245354	Li ₇ CrN ₄	123	45	0.002	0.000
40	mp-1226890	Ce ₄ H ₁₁	122	36	0.000	0.000
41	mp-695843	SbH ₇ (Br ₂ O) ₃	121	35	0.000	0.873
42	mp-1569254	Li ₂ Ni ₂ BiO ₆	120	35	0.024	0.831
43	mp-22552	Ni(AuF ₄) ₂	120	45	0.000	0.000
44	mp-1203875	Zn ₂ Te ₁₅ (H ₃ N) ₈	120	50	0.000	0.619
45	mp-1409292	MgNiF ₆	119	23	0.000	0.000
46	mp-1206799	Ba(GaH) ₂	119	32	0.000	0.000
47	mp-1203248	TiH ₂₄ C ₆ I ₃ (N ₂ O) ₆	118	33	0.030	0.000
48	mp-569603	BaCa ₄ (CuN ₂) ₂	117	37	0.000	0.219
49	mp-978854	Sr(GaH) ₂	117	30	0.000	0.000
50	mp-1200028	Al ₃ PH ₂₉ (SO ₁₄) ₂	116	26	0.029	0.062
51	mp-1104579	Ce ₂ H ₅	115	37	0.012	0.000
52	mp-643570	NiH ₂ SO ₅	115	50	0.000	0.000

Continued on next page

TABLE VI: (Continued)

	ID	Composition	Predicted T_c (K)	Dot product spread (K)	Hull energy (eV/atom)	Band gap (eV)
53	mp-1229281	$\text{Ba}_6\text{Ca}_{15}\text{Cu}_{18}\text{Hg}_3\text{O}_{43}$	114	48	0.026	0.000
54	mp-1185277	$\text{K}_8\text{Li}_{31}\text{Al}_8\text{O}_{32}$	113	26	0.015	0.000
55	mp-1229082	$\text{Ba}_6\text{Ca}_{12}\text{Cu}_{15}\text{Hg}_3\text{O}_{37}$	113	50	0.025	0.000
56	mp-720912	$\text{KZnH}_4\text{Br}_3\text{O}_2$	112	38	0.016	0.000
57	mp-570771	$\text{BaLi}_2(\text{MgSi})_2$	112	42	0.000	0.000
58	mp-1189297	Cs_3NiCl_5	112	48	0.022	0.087
59	mp-15885	Li_2US_3	111	43	0.000	0.000
60	mp-561000	$\text{Cs}_4\text{KLiFe}_2\text{F}_{12}$	111	38	0.000	0.000
61	mp-505569	CeH_2	110	46	0.017	0.000
62	mp-1228579	$\text{Ba}_2\text{Ca}_3\text{Cu}_4\text{HgO}_{10}$	110	50	0.020	0.000
63	mp-14763	Ca_3MnN_3	108	50	0.000	0.000
64	mp-1226365	$\text{Cs}_2\text{Cu}_3\text{NiF}_{10}$	107	38	0.008	0.605
65	mp-22601	$\text{Ba}_2\text{Ca}_2\text{Cu}_3\text{HgO}_8$	106	48	0.018	0.000
66	mp-989559	$\text{Cs}_2\text{NaLiF}_6$	106	48	0.000	0.000
67	mp-764501	Na_8FeO_6	105	49	0.019	0.017
68	mp-703531	$\text{CuSiH}_8(\text{O}_2\text{F}_3)_2$	105	50	0.004	0.640
69	mp-558211	$\text{Cs}_2\text{NaCoF}_6$	104	36	0.000	0.902
70	mp-1229126	$\text{Ba}_2\text{Ca}_4\text{TlCu}_5\text{O}_{13}$	104	49	0.025	0.000
71	mp-570185	$\text{Ba}_{38}\text{Na}_{58}\text{Li}_{26}\text{N}$	104	26	0.011	0.024
72	mp-755306	$\text{Li}_5\text{Ni}_3(\text{SnO}_5)_2$	103	30	0.012	0.066
73	mp-1289276	$\text{Li}_3\text{Ni}_2\text{SnO}_6$	103	30	0.030	0.344
74	mp-1292604	$\text{Li}_3\text{Ni}_2\text{SnO}_6$	103	30	0.029	0.357
75	mp-1312304	$\text{Li}_5\text{Ni}_3(\text{SnO}_5)_2$	103	30	0.011	0.619
76	mp-865625	Na_2MgSn	102	50	0.000	0.000
77	mp-29720	$\text{Li}_{21}\text{Si}_5$	101	31	0.000	0.000
78	mp-1228589	$\text{Ba}_2\text{Ca}_3\text{TlCu}_4\text{O}_{11}$	100	47	0.025	0.000
79	mp-5077	NaLi_2Sb	100	50	0.000	0.680
80	mp-14794	K_6CoS_4	100	27	0.028	0.438
81	mp-5515	Li_7MnN_4	100	43	0.000	0.656
82	mp-505213	$\text{Ba}_6\text{Na}_{16}\text{N}$	99	26	0.019	0.000
83	mp-989568	$\text{Cs}_2\text{NaMgF}_6$	99	43	0.000	0.000
84	mp-569849	$\text{Li}_{15}\text{Si}_4$	99	47	0.000	0.000
85	mp-1222798	$\text{Li}_{14}\text{MgSi}_4$	99	48	0.000	0.099
86	mp-865964	Li_2CaSn	98	35	0.000	0.000
87	mp-1246000	LiMnN_2	98	42	0.010	0.000
88	mp-1225961	CsScCuF_6	98	43	0.013	0.000
89	mp-556086	Na_5FeS_4	97	49	0.005	0.797
90	mp-697560	CoPH_6NO_5	97	46	0.008	0.707
91	mp-1203051	$\text{K}_5\text{Li}_2\text{EuF}_{10}$	97	34	0.013	0.133
92	mp-1184919	K_3Na	97	42	0.027	0.000
93	mp-1184893	K_3Na	97	42	0.027	0.000
94	mp-1184844	K_3Na	97	42	0.023	0.000
95	mp-865890	Li_2CaIn	96	37	0.000	0.000
96	mp-1229139	$\text{Ba}_{10}\text{Ca}_5\text{Cu}_{10}\text{Hg}_5\text{O}_{31}$	96	49	0.028	0.000
97	mp-1226362	$\text{Cu}_6\text{Te}_2\text{Mo}_2\text{H}_2\text{Cl}_4\text{O}_{15}$	96	39	0.029	0.000
98	mp-849362	$\text{Na}_{14}\text{Co}_2\text{O}_9$	96	47	0.009	0.892
99	mp-6879	$\text{Ba}_2\text{CaCu}_2\text{HgO}_6$	96	49	0.018	0.000
100	mp-1227794	$\text{BaSrCa}_2\text{Tl}(\text{CuO}_3)_3$	95	47	0.023	0.000

F. Materials Project, Ambient Pressure

TABLE VII: Top 100 materials in the Materials Project with the highest T_c predictions using ML model trained on ambient pressure data only, dot product spreads of less than 42 K, and energies above their convex hulls of less than 0.030 eV/atom.

	ID	Composition	Predicted T_c (K)	Dot product spread (K)	Hull energy (eV/atom)
1	mp-1205028	$H_{15}IrBr_3N_5$	189	33	0.000
2	mp-24461	$H_{12}OsN_5Cl_3O$	161	25	0.023
3	mp-1199051	$B_{10}H_{13}I$	151	40	0.000
4	mp-1199255	$B_{10}H_{13}I$	151	40	0.002
5	mp-30977	B_5H_6Br	149	40	0.000
6	mp-1197561	$B_{10}H_{13}Br$	143	39	0.000
7	mp-634446	$CsAl(H_2N)_4$	132	39	0.000
8	mp-1199374	$TaH_9N_3F_5$	132	24	0.000
9	mp-1204398	$CoMoH_{24}N_6ClO_7$	129	26	0.015
10	mp-767240	$CsPH_4(NO)_2$	129	37	0.000
11	mp-1202248	$ReH_{30}Ru_2(NCl)_{10}$	127	24	0.029
12	mp-1198073	CsH_4NF_2	125	40	0.004
13	mp-1200630	$B_8H_4PbO_{15}$	123	38	0.000
14	mp-642740	$CsLi(H_2N)_2$	123	39	0.000
15	mp-721697	$CsNa_2H_4Cl_3O_2$	122	37	0.020
16	mp-728329	$MoH_{12}Pd(NO)_4$	122	28	0.004
17	mp-733457	$CsNa_2(H_5O_3)_3$	121	39	0.003
18	mp-24135	$AlH_{18}Ru(NF)_6$	120	31	0.018
19	mp-17718	$CsKNa_2Li_{12}(SiO_4)_4$	120	37	0.000
20	mp-17125	$CsNa_3Li_{12}(GeO_4)_4$	119	35	0.001
21	mp-738629	$CsGaH_{24}(SO_{10})_2$	118	39	0.000
22	mp-570097	$Sr(Li_2P)_2$	117	36	0.000
23	mp-23835	$HgH_{12}C_2(Br_2N_3)_2$	117	27	0.000
24	mp-15845	$SrLi_4N_2$	116	37	0.000
25	mp-643359	$Rb_2Sn(H_2N)_6$	115	40	0.028
26	mp-1190754	$Cd(BH_4)_2$	115	36	0.000
27	mp-1191186	$CsAlH_4(OF_2)_2$	114	31	0.000
28	mp-28204	CsH_7O_4	113	41	0.000
29	mp-722455	$RbLi_2(H_2N)_3$	112	34	0.006
30	mp-1179724	$RbLi_2(H_2N)_3$	112	34	0.006
31	mp-755322	$BaNa_6O_4$	112	40	0.018
32	mp-1179387	$ReH_{12}N_3Cl_4O_3$	111	24	0.001
33	mp-1193778	$RbAl(H_2N)_4$	111	38	0.000
34	mp-1251539	$RbAl(H_2N)_4$	111	38	0.002
35	mp-1202291	$BiB_4(HO_3)_3$	109	41	0.008
36	mp-643371	$K_2Sn(H_2N)_6$	109	39	0.000
37	mp-643905	$Sr(H_2N)_2$	108	39	0.000
38	mp-8611	Li_8CeO_6	108	41	0.002
39	mp-556177	$Cs_2Cu_2Si_8O_{19}$	108	40	0.000
40	mp-696969	$Sr(H_2N)_2$	108	39	0.001
41	mp-707501	$CdH_{24}C_4(BrN_2)_6$	108	21	0.002
42	mp-24523	$CsAlH_{24}(SeO_{10})_2$	107	30	0.000
43	mp-1212424	$HgH_{12}C_2(I_2N_3)_2$	107	33	0.000
44	mp-1198273	$SrP_{11}H_{28}N_9$	107	40	0.000
45	mp-604315	$Zn(BH_4)_2$	106	36	0.029
46	mp-1193046	$CsMgAs(H_6O_5)_2$	106	39	0.000
47	mp-1191586	$Zn(BH_4)_2$	106	36	0.000
48	mp-1201403	$ZnH_{13}RuN_5(Cl_2O)_2$	106	34	0.018
49	mp-510073	$RbLi(H_2N)_2$	105	35	0.000
50	mp-1198095	$CsAlH_{24}(SO_{10})_2$	104	30	0.000
51	mp-696329	$RbCa(H_2N)_3$	104	39	0.000

Continued on next page

TABLE VII: (Continued)

	ID	Composition	Predicted T_c (K)	Dot product spread (K)	Hull energy (eV/atom)
52	mp-1016214	$\text{Bi}_2\text{H}_{18}\text{C}_3(\text{IN})_9$	103	38	0.000
53	mp-1203928	$\text{NbH}_8\text{N}_2\text{OF}_5$	103	27	0.000
54	mp-553901	$\text{H}_6\text{RuN}_3\text{Cl}_3\text{O}$	103	33	0.009
55	mp-974267	KH_8N_3	103	39	0.000
56	mp-695172	$\text{CsMg}_{12}\text{Al}_{25}\text{Si}_{29}\text{O}_{108}$	103	39	0.022
57	mp-695133	$\text{CsMg}_4\text{Al}_9(\text{SiO}_4)_9$	102	39	0.019
58	mp-8407	Li_3LaP_2	102	41	0.000
59	mp-574928	$\text{CsPH}_3\text{O}_3\text{F}$	101	38	0.011
60	mp-1226440	$\text{CsAlSi}_5\text{O}_{12}$	101	40	0.015
61	mp-973335	$\text{Hg}_2\text{H}_6\text{CN}_3\text{Cl}_5$	101	33	0.001
62	mp-707900	$\text{PrH}_{20}\text{S}_4\text{N}_5\text{O}_{16}$	100	36	0.009
63	mp-707956	$\text{KLi}_7(\text{H}_2\text{N})_8$	100	32	0.022
64	mp-695891	$\text{CdH}_6\text{C}(\text{BrN})_3$	100	26	0.000
65	mp-1205148	$\text{PrH}_{20}\text{S}_4\text{N}_5\text{O}_{16}$	100	36	0.005
66	mp-505791	$\text{ReH}_8(\text{Br}_3\text{N})_2$	100	37	0.000
67	mp-1204099	$\text{BiH}_{14}\text{C}_4\text{N}_8\text{Cl}_5\text{O}_2$	99	21	0.001
68	mp-632724	$\text{ReH}_8(\text{NCl}_3)_2$	99	27	0.000
69	mp-2646923	$\text{Lu}(\text{BH}_4)_3$	99	35	0.000
70	mp-2646940	$\text{Lu}(\text{BH}_4)_3$	99	35	0.003
71	mp-1220119	$\text{PrH}_{18}(\text{BrO}_6)_3$	99	39	0.000
72	mp-1201394	$\text{K}_2\text{Zn}(\text{H}_2\text{N})_4$	99	28	0.004
73	mp-562920	$\text{CsAl}(\text{SiO}_3)_2$	99	41	0.000
74	mp-722979	$\text{K}_2\text{Zn}(\text{H}_2\text{N})_4$	99	28	0.000
75	mp-1195558	$\text{Zn}(\text{H}_{15}\text{N}_8)_2$	98	27	0.000
76	mp-1215828	$\text{Zn}_8\text{B}_4\text{H}_3\text{O}_{15}\text{F}$	98	41	0.005
77	mp-1211490	$\text{KMnH}_8(\text{NF}_3)_2$	98	29	0.028
78	mp-1195877	$\text{B}_{10}\text{H}_{13}\text{Cl}$	98	39	0.001
79	mp-1198991	$\text{Rb}_3\text{H}_{12}\text{N}_5$	98	39	0.000
80	mp-569827	$\text{HgH}_8\text{C}_2\text{Br}_3\text{N}$	98	32	0.019
81	mp-676956	$\text{Zn}_8\text{B}_4\text{H}_3\text{O}_{15}\text{F}$	98	41	0.017
82	mp-1212542	$\text{H}_9\text{C}_3\text{IN}_6\text{O}$	97	18	0.012
83	mp-866716	$\text{RbLiH}_{12}\text{Se}_3\text{N}_4$	97	39	0.000
84	mp-1196579	$\text{Cs}_2\text{CuSi}_5\text{O}_{12}$	97	41	0.006
85	mp-697558	$\text{SbH}_{12}\text{C}_2\text{N}_6\text{F}_5$	97	18	0.000
86	mp-761178	$\text{CsMgP}(\text{H}_6\text{O}_5)_2$	97	37	0.000
87	mp-758775	$\text{CsMnH}_4(\text{OF}_2)_2$	97	39	0.000
88	mp-1198673	$\text{SnH}_{12}\text{C}_2(\text{NCl})_6$	96	21	0.000
89	mp-2646967	$\text{La}(\text{BH}_4)_3$	96	36	0.000
90	mp-17240	$\text{RbNa}_3\text{Li}_{12}(\text{SiO}_4)_4$	96	22	0.000
91	mp-1199487	$\text{MoH}_8(\text{S}_2\text{N})_2$	96	29	0.000
92	mp-1196086	$\text{CaH}_{24}(\text{BrN}_4)_2$	96	31	0.000
93	mp-761870	$\text{H}_7\text{C}_3\text{N}_6\text{Cl}$	95	18	0.000
94	mp-697905	$\text{LaMg}_2\text{NiH}_7$	95	38	0.005
95	mp-1199443	$\text{SnH}_6\text{C}(\text{NCl})_3$	95	22	0.000
96	mp-721171	$\text{CsMgH}_{12}(\text{ClO}_2)_3$	95	38	0.000
97	mp-554483	$\text{KCd}_3\text{H}_8\text{Cl}_7\text{O}_4$	95	39	0.014
98	mp-1201495	$\text{Cs}_2\text{Al}_2\text{Si}_3\text{O}_{10}$	94	41	0.011
99	mp-653631	$\text{BaNa}_4\text{Cu}_3\text{F}_{12}$	94	40	0.014
100	mp-722866	$\text{SbH}_6\text{CN}_3\text{F}_4$	94	20	0.000

G. Materials Project, Ambient Pressure, Small Band Gaps

TABLE VIII: Top 100 materials in the Materials Project with the highest T_c predictions using ML model trained on ambient pressure data only, dot product spreads of less than 42 K, and energies above their convex hulls of less than 0.030 eV/atom, and band gaps of less than 1.000 eV.

	ID	Composition	Predicted T_c (K)	Dot product spread (K)	Hull energy (eV/atom)	Band gap (eV)
1	mp-1202248	ReH ₃₀ Ru ₂ (NCl) ₁₀	127	24	0.029	0.031
2	mp-24135	AlH ₁₈ Ru(NF) ₆	120	31	0.018	0.000
3	mp-570097	Sr(Li ₂ P) ₂	117	36	0.000	0.976
4	mp-15845	SrLi ₄ N ₂	116	37	0.000	0.887
5	mp-556177	Cs ₂ Cu ₂ Si ₈ O ₁₉	108	40	0.000	0.755
6	mp-8407	Li ₃ LaP ₂	102	41	0.000	0.612
7	mp-505791	ReH ₈ (Br ₃ N) ₂	100	37	0.000	0.813
8	mp-866716	RbLiH ₁₂ Se ₃ N ₄	97	39	0.000	0.801
9	mp-1196579	Cs ₂ CuSi ₅ O ₁₂	97	41	0.006	0.180
10	mp-697905	LaMg ₂ NiH ₇	95	38	0.005	0.951
11	mp-653631	BaNa ₄ Cu ₃ F ₁₂	94	40	0.014	0.160
12	mp-2646929	Ce(BH ₄) ₃	90	38	0.000	0.344
13	mp-1777	Li ₁₅ Ge ₄	89	41	0.000	0.000
14	mp-18713	Ba ₂ Li(CuO ₂) ₃	88	41	0.008	0.000
15	mp-1218762	Sr ₂ Li ₇ CuN ₄	88	27	0.000	0.568
16	mp-28994	KLi ₂ As	87	34	0.000	0.688
17	mp-1226362	Cu ₆ Te ₂ Mo ₂ H ₂ Cl ₄ O ₁₅	87	37	0.029	0.000
18	mp-29630	Li ₇ Ge ₂	87	41	0.011	0.000
19	mp-865964	Li ₂ CaSn	86	35	0.000	0.000
20	mp-736	Li ₃ P	85	37	0.000	0.700
21	mp-867342	Li ₃ Ge	84	39	0.009	0.000
22	mp-638252	Cs ₂ Cu ₃ (P ₂ O ₇) ₂	84	40	0.000	0.755
23	mp-505089	Li ₄ TbF ₈	83	29	0.012	0.000
24	mp-2251	Li ₃ N	83	36	0.000	0.984
25	mp-849362	Na ₁₄ Co ₂ O ₉	83	34	0.009	0.892
26	mp-6127	BaCu(Si ₂ O ₅) ₂	82	40	0.013	0.855
27	mp-1203248	TiH ₂₄ C ₆ I ₃ (N ₂ O) ₆	82	19	0.030	0.000
28	mp-6496	Ba ₂ Na(CuO ₂) ₃	82	41	0.000	0.000
29	mp-759174	Na ₆ CoO ₄	82	33	0.019	0.790
30	mp-14885	Rb ₂ Cu ₂ Si ₈ O ₁₉	81	39	0.002	0.659
31	mp-757	Li ₃ As	81	40	0.000	0.636
32	mp-27454	BaLi ₂ Si	81	31	0.000	0.000
33	mp-1185028	Li ₂ CaAl	81	19	0.020	0.000
34	mp-1195515	K ₂ U(Si ₂ O ₅) ₃	81	41	0.008	0.016
35	mp-867171	SrLi ₂ Sn	81	38	0.000	0.000
36	mp-1194845	K ₂ U(Si ₂ O ₅) ₃	81	41	0.009	0.116
37	mp-865890	Li ₂ CaIn	81	35	0.000	0.000
38	mp-1228881	CsMnCuF ₆	80	34	0.009	0.000
39	mp-20067	Na ₅ OsO ₆	80	32	0.000	0.000
40	mp-505786	CeH ₁₄ Cl ₃ O ₇	79	41	0.011	0.023
41	mp-8406	Li ₃ NdAs ₂	79	38	0.000	0.036
42	mp-1204167	CeH ₁₄ Cl ₃ O ₇	79	41	0.014	0.328
43	mp-1194910	Mn(BH ₄) ₂	78	33	0.013	0.000
44	mp-698125	NpH ₁₆ C ₄ N ₈ ClO ₆	78	20	0.010	0.429
45	mp-1038816	CeMg	78	41	0.016	0.000
46	mp-865986	Li ₂ CaGe	77	22	0.003	0.101
47	mp-1182662	BaCu(SiO ₃) ₂	77	41	0.022	0.613
48	mp-555332	BaCu(SiO ₃) ₂	77	41	0.021	0.000
49	mp-1197716	BaCu(SiO ₃) ₂	77	41	0.011	0.640
50	mp-1200513	BaCu ₂ Ge ₃ (HO ₅) ₂	77	41	0.021	0.579
51	mp-676620	Li ₂₇ As ₁₀	77	39	0.025	0.000

Continued on next page

TABLE VIII: (Continued)

	ID	Composition	Predicted T_c (K)	Dot product spread (K)	Hull energy (eV/atom)	Band gap (eV)
52	mp-1198097	$\text{Sr}_2\text{SnH}_{18}(\text{SeO}_3)_4$	77	38	0.000	0.819
53	mp-764501	Na_8FeO_6	76	34	0.019	0.017
54	mp-1598	Na_3P	76	38	0.000	0.403
55	mp-1209376	RbAlCuF_6	76	40	0.000	0.000
56	mp-19026	Na_4FeO_3	76	32	0.015	0.763
57	mp-1226046	$\text{CsInCu}(\text{PO}_3)_6$	76	40	0.000	0.590
58	mp-1196302	$\text{MnTe}_4(\text{H}_3\text{N})_6$	76	41	0.002	0.142
59	mp-764233	Na_8CoO_6	76	34	0.000	0.951
60	mp-5077	NaLi_2Sb	76	36	0.000	0.680
61	mp-1187184	SrLi_2In	76	40	0.025	0.000
62	mp-8469	Li_8IrO_6	76	29	0.005	0.000
63	mp-3988	BaCu_3O_4	75	41	0.014	0.000
64	mp-27932	Li_9Ge_4	74	37	0.003	0.000
65	mp-2362521	Li_9Ge_4	74	37	0.000	0.000
66	mp-1203793	$\text{CeB}_4\text{H}_2\text{ClO}_8$	74	41	0.000	0.450
67	mp-697156	$\text{BaCu}_2\text{P}_2\text{H}_2\text{O}_9$	74	40	0.009	0.008
68	mp-1202929	$\text{AgH}_8\text{C}_2\text{S}_2\text{N}_4\text{Cl}$	74	19	0.024	0.000
69	mp-1227989	$\text{BaSrAl}_{20}(\text{CuO}_{17})_2$	74	40	0.017	0.000
70	mp-1194032	$\text{Pu}(\text{SO}_6)_2$	73	39	0.019	0.000
71	mp-6960	Rb_2SiAs_2	73	26	0.000	0.955
72	mp-1196467	$\text{Rb}_2\text{Cu}_3\text{P}_4(\text{O}_3\text{F})_4$	71	35	0.000	0.527
73	mp-1228431	$\text{Ba}_2\text{Cu}_3\text{BrClO}_4$	71	39	0.002	0.000
74	mp-570771	$\text{BaLi}_2(\text{MgSi})_2$	71	32	0.000	0.000
75	mp-1226351	$\text{CsMn}_2(\text{PO}_3)_6$	71	41	0.000	0.000
76	mp-1225984	$\text{CsMn}_2(\text{PO}_3)_6$	71	41	0.002	0.000
77	mp-531553	Li_5GeN_3	71	33	0.000	0.001
78	mp-654359	$\text{BaCu}_2(\text{PO}_4)_2$	71	41	0.012	0.326
79	mp-1213081	$\text{CsMn}_2(\text{PO}_3)_6$	71	41	0.003	0.000
80	mp-1223676	$\text{K}_2\text{Na}_4\text{Co}_2\text{O}_5$	70	37	0.000	0.076
81	mp-1207213	K_2LiCuF_6	70	36	0.000	0.925
82	mp-1147739	$\text{Ba}(\text{CuO})_4$	70	39	0.026	0.000
83	mp-1018766	Li_3LaAs_2	70	40	0.000	0.502
84	mp-756356	Li_8FeO_6	70	30	0.029	0.000
85	mp-763305	Li_8FeO_6	70	30	0.025	0.000
86	mp-504863	KNa_2FeO_3	69	37	0.010	0.502
87	mp-653840	$\text{KCu}_7\text{TeS}_5\text{ClO}_{24}$	69	40	0.000	0.000
88	mp-1186135	NaLi_2Sn	69	38	0.025	0.000
89	mp-1177487	$\text{Li}_{47}(\text{CoO}_4)_8$	69	39	0.023	0.828
90	mp-542450	$\text{Cs}_3\text{Mn}_4(\text{PO}_3)_{12}$	69	41	0.000	0.000
91	mp-1218363	$\text{SrCa}(\text{CuO}_2)_2$	68	36	0.027	0.000
92	mp-23129	$\text{Ba}_2\text{Cu}_3(\text{BrO}_2)_2$	68	39	0.000	0.000
93	mp-1178147	K_3TlPCO_7	68	33	0.021	0.000
94	mp-707454	$\text{Li}(\text{H}_3\text{N})_4$	68	23	0.028	0.000
95	mp-1218417	$\text{SrCa}(\text{CuO}_2)_2$	68	36	0.026	0.000
96	mp-976023	Li_3Ga	68	38	0.021	0.000
97	mp-1147551	KCuTeO_6	68	41	0.001	0.780
98	mp-551283	$\text{Na}_2\text{Cu}_2\text{TeO}_6$	68	38	0.005	0.066
99	mp-18124	KAlCuF_6	67	39	0.003	0.081
100	mp-1196923	Li_2TbF_6	67	31	0.015	0.000