then it could possibly be formable. We acknowledge that whether a given stoichiometry will form a compound or not will depend on synthesis conditions, which we cannot predict. Out of the 235 compounds predicted by ML to be in a

expected properties of materials not yet known to exist. Since the input to the learning process consists of previously synthesized compounds, the ensuing ML predictions provide insights into chemistries in a given stoichiometry that are probable in a crystal structure type without commenting on whether they are thermodynamically stable or not.

B. Predicting new ABO_3 perovskites by ML

Our overarching ML strategy is shown in Fig. 1. We first demonstrate that the ML models can classify the known 390 ABO_3 compounds into the 254 perovskites and 136 that are not perovskites with 90% average cross-validation (CV) accuracy determined by a stratified CV procedure [Fig. 1(a)]. This success, in part, requires identifying effective chemical descriptors (features) that enable such classification. Then, we build another set of ML models to classify all formable 254 perovskite structures into the 22 known cubic perovskites and 232 known noncubic perovskites finding similarly 94% CV accuracy [Fig. 1(b)]. We interpret that the misclassifications of our models as more of a source of important physical information than as a failing of the model. For example, $KTaO_3$ and $SrVO_3$ were classified as noncubic perovskite by ML whereas they were reported to be formed as cubic perovskites. They are likely poised to undergo a structural transition from the experimentally observed cubic to noncubic perovskite structure. This possibility remains to be experimentally validated (discussed in Sec. IV C).

We next apply the trained ML models constructed from

III. INPUTS AND METHODS

A. Database of known ABO_3 compounds

Our database of ABO_3 compounds consists of 390 compounds and was created via an augmentation of the database of 354 ABO_3 compounds explored earlier by Pilania *et al.* [26]. These data included those compiled by Zhang *et al.* [5] who gathered their data from a number of resources, including the Inorganic Crystal Structure Database (ICSD) and other published data. We added to our earlier 354 compounds 36 new ABO_3 compounds taken from Ref. [37] and those compiled by Emery *et al.* [15]. We note that in all 390 compounds the sum of the valences of A and B adds to six so these are "charge balanced compounds." For example, each A-B pair has nominal I-V, II-IV, or III-III valences. No A-B pairs in this set have IV-II or V-I valences. Each previously documented ABO_3 compound has a label signifying whether it is a perovskite or not. Compounds with the ABO

tree boosting, we set the subsampling of the training data at 50%, the number of ensembles at 2500 for the perovskite or not case and 2000 for the cubic or not case, and the learning rate at 0.001. In the RFC case, deep tree depths resulted in a signiÞcant overÞtting of the training data, often approaching 100% accuracy but with a large variance. We simply decreased the depth, observing the accuracy of the predictions on the test data increasing and the variance decreasing. When the mean accuracy started to decrease, we stopped. We adjusted the hyperparameters for GTBC similarly but set the maximum tree depth of its trees to 3 to make the classiÞer weak. We made these adjustments for the octahedral and tolerance factor case, whose model gave the initial highest accuracy and hence had the greatest likelihood being overÞt, and applied them to the classiÞers for the other feature pair cases.

In Table I, we give the mean and standard deviations of

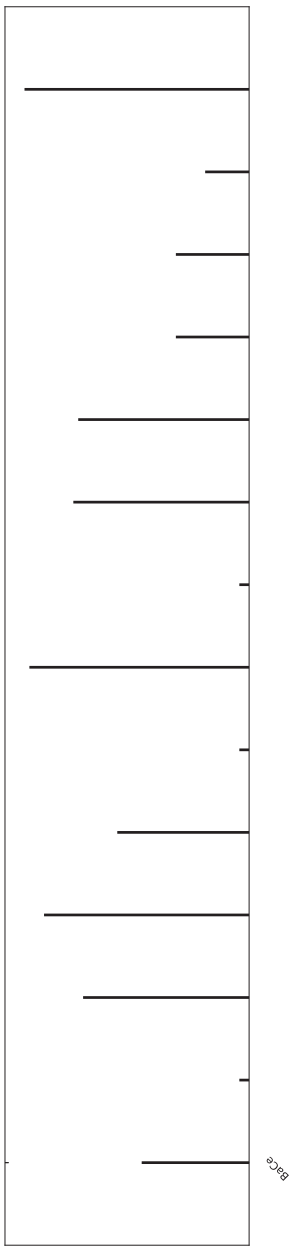
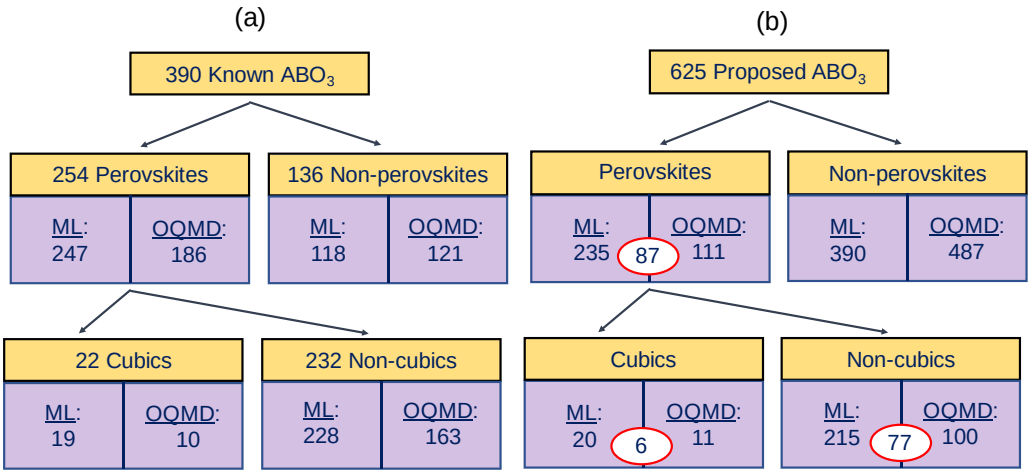

BaCe

TABLE V. Comparison of the classiÞcations of experimentally synthesized ABO_3 compounds (referred to as ÒDATAÓ in the table) with the stability predicted from the OQMD database. Out of the 390 compounds in our database, the stability of 387 were computed in OQMD. The word ÒcubicsÓ refers to a perovskite in the cubic structure ($Pm\bar{3}m$).


(a)
390 Known ABO3
254 Perovskites
ML: 247
OQMD: 186
136 Non-perovskites
ML: 118
OQMD: 121
22 Cubics
ML: 19
OQMD: 10
232 Non-cubics
ML: 228
OQMD: 163
(b)
625 Proposed ABO3
Perovskites
ML: 235
87
OQMD: 111
Non-perovskites
ML: 390
OQMD: 487
Cubics
ML: 20
6
OQMD: 11
Non-cubics
ML: 215
77
OQMD: 100

we were mainly targeting consistent performance for the cubic or not case which was difficult to achieve because so few cubics are in the data. In particular, we found that other cross-validation techniques were giving predictions that sometimes had large variances most likely due to overfitting the training data.

We emphasize our ML analyses and predictions are statistical in nature and hence are always subject to changes caused by fluctuations. Further, other ML approaches might produce results with higher accuracy if they were to use more features and optimize their hyperparameters for each case considered as opposed to our selecting just two features and using a one-size-fits-all setting of the hyperparameters. In another paper [26], for example, we demonstrated that using more than just pairs from the set of four feature pairs we could increase the accuracy of the predictions to nearly 95%. However, similar predictions of possible new perovskites were still made. In part, this improved accuracy is likely a consequence of using more parameters to fit the data as opposed to using features that delineate trends in the data better. Increasing the number of

(continued.)

Formula	ML prediction	OQMD prediction	H	E	SG
$LuNiO_3$	P	P	Š2.44	37	62
$NdCuO_3$	P	P	Š2.18	74	62
$NdRuO_3$	P	P	Š2.34	41	62
$PbPaO_3$	P	P	Š2.41	Š22	62
$PbPuO_3$	P	P	Š2.54	Š55	62
$PrCuO_3$	P	P	Š2.17	3	167
$PrInO_3$	P	P	Š2.71	5	62
$PuGaO_3$	P	P	Š2.9	Š28	62
$SmCuO_3$	P	P	Š2.22	47	62
$SmGaO_3$	P	P	Š2.92	6	62
$SmRuO_3$	P	P	Š2.37	49	62
$SrCrO_3$	P	P	Š2.56	41	62
$SrNpO_3$	P	P	Š3.42	Š14	62
$SrPaO_3$	P	P	Š3.18	Š144	62
$SrUO_3$	P	P	Š3.49	Š18	62

(continued.)

Formula	ML prediction	OQMD prediction	H	E	SG
$TbCuO_3$	P	P	Š2.14	Š40	62
$TbGaO_3$	P	P	Š2.83	15	62
$TbNiO_3$	P	P	Š2.27	8	62
$TbScO_3$	P	P	Š3.66	19	62
$TlMnO_3$	P	P	Š1.43	51	62
$TmCoO_3$	P	P	Š2.57	27	62
$TmGaO_3$	P	P	Š3	19	62
$YbCoO_3$	P	P	Š2.11	Š79	62
$YbRhO_3$	P	P	Š2.11	Š89	62
$YbRuO_3$	P	P	Š2.25	Š83	62
$YbScO_3$	P	P	Š3.21	98	62
$EuErO_3$	P	P	Š3.21	98	62
$EuLuO_3$	P	P	Š3.26	94	62
$EuTmO_3$	P	P	Š3.23	90	62

[1] A. S. Bhalla, R. Guo, and R. Roy, Mater. Res. Innov. 4, 3 (2000).
[2] J. B. Goodenough, Rep. Prog. Phys. 67, 1915 (2004).
[3] M. A. Pe–a and J. L. G. Fierro, Chem. Rev. Rep Rev.

[37] K. Ito, K. Tezuka, and Y. Hinatsu, J. Solid State Chem. 157, 173 (2001).
[38] C.-Q. Jin, J.-S. Zhou, J. B. Goodenough, Q. Q. Liu, J. G. Zhao, L. X. Yang, Y. Yu, R. C. Yu, T. Katsura, A. Shatskiy, and E. Ito, Proc. Natl. Acad. Sci. U. S. A. 105, 7115 (2008).
[39] D. M. Giaquinta and H.-C. zur Loye, Chem. Mater. 6, 365 (1994).
[40] L. M. Feng, L. Q. Jiang, M. Zhu, H. B. Liu, X. Zhou, and C. H. Li, J. Phys. Chem. Solids 69, 967 (2008).
[41] C. Li, K. C. K. Soh, and P. Wu, J. Alloys Compd. 372, 40 (2004).
[42] R. D. Shannon, Acta. Cryst. A 32, 751 (1976).
[43] I. D. Brown, Chem. Rev. 109, 6858 (2009).
[44] P. Villars, K. Cenzual, J. Daams, Y. Chen, and S. Iwata, J. Alloys Compd. 367, 167 (2004).
[45] L. Breiman, Mach. Learn. 45, 5 (2001).
[46] J. H. Friedman, Ann. Stat. 29, 1189 (2001).
[47] 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, J. Mach. Learn. Res. 12, 2825 (2011).
[48] G. Kresse and J. FurthmŸller, Comput. Mater. Sci. 6, 15 (1996).
[49] G. Kresse and J. FurthmŸller, Phys. Rev. B 54, 11169 (1996).
[50] G. Kresse and D. Joubert, Phys. Rev. B 59, 1758 (1999).
[51] J. P. Perdew, K. Burke, and M. Ernzerhof, Phys. Rev. Lett. 77, 3865 (1996).
[52] S. L. Dudarev, G. A. Botton, S. Y. Savrasov, C. J. Humphreys, and A. P. Sutton, Phys. Rev. B 57, 1505 (1998).
[53] L. Wang, T. Maxisch, and G. Ceder, Phys. Rev. B 73, 195107 (2006).
[54] A. A. Emery and C. Wolverton, Sci. Data 4, 170153 (2017).
[55] P. Giannozzi, S. Baroni, N. Bonini, M. Calandra, R. Car, C. Cavazzoni, D. Ceresoli, G. L. Chiarotti, M. Cococcioni, I. Dabo, A. D. Corso, S. de Gironcoli, S. Fabris, G. Fratesi, R. Gebauer, U. Gerstmann, C. Gougoussis, A. Kokalj, M. Lazzeri, L. Martin-Samos, N. Marzari, F. Mauri, R. Mazzarello, S. Paolini, A. Pasquarello, L. Paulatto, C. Sbraccia, S. Scandolo, G. Sclauzero, A. P. Seitsonen, A. Smogunov, P. Umari, and R. M. Wentzcovitch, J. Phys.: Condens. Matter 21, 395502 (2009).
[56] J. P. Perdew, A. Ruzsinszky, G. I. Csonka, O. A. Vydrov, G. E. Scuseria, L. A. Constantin, X. Zhou, and K. Burke, Phys. Rev. Lett. 100, 136406 (2008).
[57] D. Vanderbilt, Phys. Rev. B 41, 7892 (1990).
[58] A. D. Corso, Comput. Mater. Sci. 95, 337 (2014).
[59] M. Retuerto, S. Skiadopoulou, M.-R. Li, A. M. Abakumov, M. Croft, A. Ignatov, T. Sarkar, B. M. Abbett, J. Pokorn , M. Savinov, D. Nuzhnyy, J. Prokleška, M. Abeykoon, P. W. Stephens, J. P. Hodges, P. Yan, C. J. Fennie, K. M. Rabe, S. Kamba, and M. Greenblatt, Inorg. Chem. 55, 4320 (2016).
[60] M. De La Pierre, R. Orlando, L. Maschio, K. Doll, P. Ugliengo, and R. Dovesi, J. Comput. Chem. 32, 1775 (2011).
[61] C. Eames, J. M. Frost, P. R. F. JF1 1 Tr[61] R. F FurthmŸller,