

ARTICLE OPEN



Machine learning and evolutionary prediction of superhard B-C-N compounds

Wei-Chih Chen¹, Joanna N. Schmidt¹, Da Yan², Yogesh K. Vohra¹ and Cheng-Chien Chen¹✉

We build random forests models to predict elastic properties and mechanical hardness of a compound, using only its chemical formula as input. The model training uses over 10,000 target compounds and 60 features based on stoichiometric attributes, elemental properties, orbital occupations, and ionic bonding levels. Using the models, we construct triangular graphs for B-C-N compounds to map out their bulk and shear moduli, as well as hardness values. The graphs indicate that a 1:1 B-N ratio can lead to various superhard compositions. We also validate the machine learning results by evolutionary structure prediction and density functional theory. Our study shows that BC₁₀N, B₄C₅N₃, and B₂C₃N exhibit dynamically stable phases with hardness values >40 GPa, which are superhard materials that potentially could be synthesized by low-temperature plasma methods.

npj Computational Materials (2021)7:114; <https://doi.org/10.1038/s41524-021-00585-7>

INTRODUCTION

Superhard materials exhibit a Vickers hardness $H \geq 40$ GPa, and they have extensive applications such as abrasives, cutting tools, and protective coatings^{1–5}. Diamond is the hardest material (with $H \sim 100$ GPa), but its applications are limited by size and cost⁶. It is also not suitable for oxidizing conditions or high-speed machining of ferrous alloys because of chemical reactions with iron group elements^{6,7}. One promising class of superhard materials involve light elements B, C, N, and O⁸. These elements can form multiple short covalent bonds, which make the crystal structure difficult to break apart. One notable example is cubic boron nitride (c-BN), which has a reported hardness between $H \sim 50$ –70 GPa⁹. Other examples are boron carbides, ranging from weakly boron-incorporated diamond structures like BC₅^{10–12} to boron-rich B₁₂ icosahedron-intercalated structures like B₅₀C₂^{13–15}. For ternary compounds, several superhard B-C-N compositions have been reported^{16,17}, such as BC₂N^{18,19} and BC₄N¹⁹. Other superhard B-C-O^{20,21}, B-N-O^{22,23}, and C-N-O²⁴ compounds also have been studied. However, due to the huge phase space of possible element combinations, it remains challenging to explore ternary superhard materials.

First-principles simulations based on density functional theory (DFT) have played important roles in predicting superhard compounds. However, ab initio methods are still computationally expensive and size-limited. On the other hand, data-driven approaches have proven to be powerful and efficient in exploring materials^{25–30}—thanks to recent advance in computing hardwares, development in machine learning algorithms, and availability of online materials database. For example, Meredig et al.³¹ have constructed a machine learning model to screen over 1.6 million ternary compositions and predicted 4500 potentially stable ternary materials. Therefore, data-driven machine learning approaches are promising for large-scale materials design and discovery.

In principle, a machine learning framework can be implemented with different material features or descriptors for a wide range of target properties. Two popular properties to predict are bulk and shear moduli^{32–37}, which are also correlated with the material

hardness. For example, de Jong et al.³³ developed a technique based on gradient boosting and used features like the volume per atom and cohesive energy. Mansouri et al.³⁶ used support vector machines and combined elemental and structural properties as descriptors, where the cohesive energy was also identified as a crucial feature. These machine learning studies typically can achieve high prediction accuracy with only a few thousands of training data points. However, using cohesive energy, volume, melting point, crystal symmetry, and so on as features may be less ideal, as obtaining these information for new compounds would require additional measurements or calculations.

In this paper, we develop random forests models to predict material mechanical properties, by using only features that can be derived directly from the chemical formula. The resulting machine learning models thereby can achieve large-scale prediction of superhard and ultraincompressible materials for extreme environment applications. We also employ evolutionary structure prediction and DFT calculations to further validate the machine learning results. In particular, we propose three superhard compositions—BC₁₀N, B₄C₅N₃, and B₂C₃N—and fully characterize their structural, phonon, and electronic properties using first-principles calculations. These superhard compounds are all dynamically stable with relatively low formation energy, so they can potentially be synthesized by low-temperature plasma methods, without the need of high-temperature high-pressure conditions. It is noted that BC₁₀N has a computed hardness ~ 87 GPa close to that of diamond, and it is expected to show better performance in higher temperature and humidity environments. Our computational flowchart is summarized in Fig. 1 and discussed in detail in the “Methods” section.

RESULTS AND DISCUSSION

Machine learning prediction of mechanical properties

Figure 2a shows histograms of the bulk and shear moduli computed by DFT (denoted, respectively, as K_{DFT} and G_{DFT}) for 10,421 samples acquired from the Materials Project database³⁸. Here the DFT modulus values represent the Voigt–Reuss–Hill

¹Department of Physics, University of Alabama at Birmingham, Birmingham, AL, USA. ²Department of Computer Science, University of Alabama at Birmingham, Birmingham, AL, USA. ✉email: chencc@uab.edu

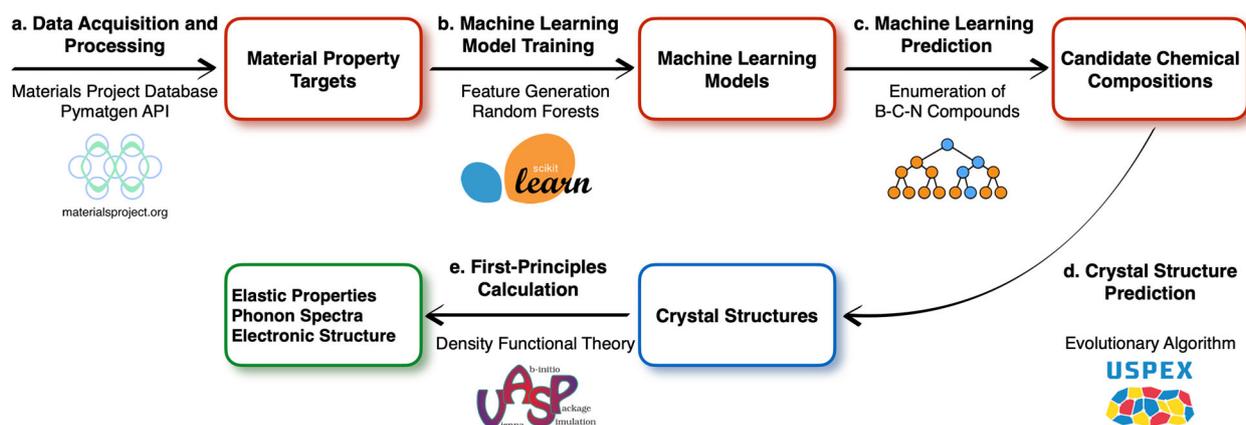


Fig. 1 Computational flowchart of data-driven discovery of superhard materials. **a** Data acquisition and processing using the Materials Project³⁸ database and its application programming interface (API) Pymatgen⁶⁰. **b** Machine learning model training with handcrafted features and regression algorithms implemented in the SCIKIT-LEARN library⁶⁵. **c** Random forests prediction of chemical compositions for candidate superhard materials. **d** Crystal structure prediction of a given chemical formula using evolutionary algorithms implemented in the USPEX program^{76–78}. **e** First-principles validation of the machine learning results with density functional theory calculations using the VASP software^{79,80}.

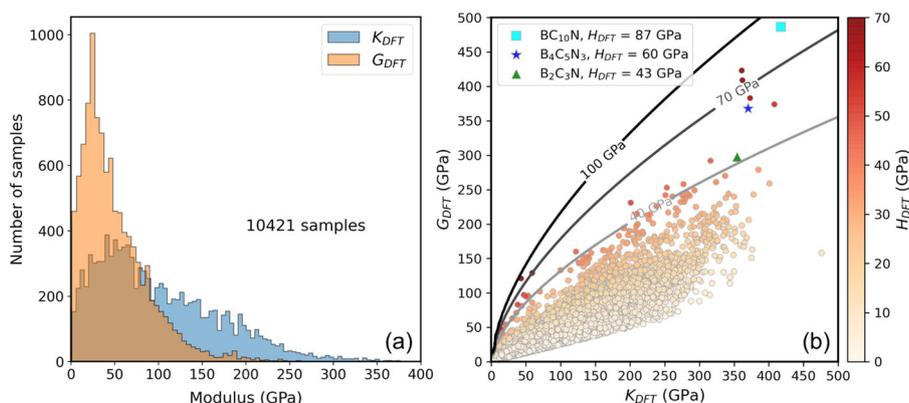


Fig. 2 Target data acquired from the Materials Project database³⁸. **a** Histogram and **b** scatter plot of bulk (K) and shear (G) moduli for 10,421 samples based on density functional theory (DFT) calculations. In **b**, the false-color intensity represents the Vickers hardness (H) computed by Tian's empirical model using K and G as inputs. The solid curves represent the hardness contours using Tian's model⁴³. These contour lines can help quickly locate compounds with superhardness, with the caveat that the model's applicability might be more limited in the low-bulk/low-shear modulus region. The three proposed superhard compounds BC_{10}N , $\text{B}_4\text{C}_5\text{N}_3$, and $\text{B}_2\text{C}_3\text{N}$ are highlighted respectively by the \square , \star , and \triangle symbols.

average moduli^{39–41}, and the medians of K_{DFT} and G_{DFT} in Fig. 2a are 84 and 40 GPa, respectively. Since the accuracy and applicability of a machine learning model largely depend on the training data, we have set a few criteria to select suitable samples during data acquisition.

First, we have excluded sample materials with a formation energy ≥ 0.2 eV per atom, as they are thermodynamically unfavorable. Second, we have neglected samples whose Voigt and Reuss modulus values differ by >50 GPa; this class of samples are typically layered quasi-two-dimensional materials, like graphite or hexagonal BN, which are not the focus of our study. Third, we have utilized the Pugh's ratio $k(\equiv G/K)$ ⁴² to further filter out materials with $k < 0.25$ due to their extremely small hardness, as well as materials with $k > 4.0$, which represents an extreme high-hardness structure (with $H > 200$ GPa) usually computed under high pressure. With these selection criteria, there are in total 10,421 samples considered in our machine learning study.

Figure 2b shows the scatter plot for the distribution of bulk and shear moduli. The false-color intensity represents the corresponding material hardness (H), which is calculated by using Tian's

empirical model⁴³:

$$H = 0.92k^{1.137}G^{0.708} \quad (1)$$

This empirical formula dictates that superhardness requires a large Pugh's ratio k and/or a high shear modulus G . Using Tian's model, we also plot hardness contour lines in Fig. 2b. Materials in the contour region between $H = 40$ and $H = 100$ GPa are mostly compounds like C, BC_2N , c-BN, and M_xB_y (M: Be or transition metal). The three proposed superhard ternary compounds— BC_{10}N , $\text{B}_4\text{C}_5\text{N}_3$, and $\text{B}_2\text{C}_3\text{N}$ —are highlighted, respectively, by the \square , \star , and \triangle symbols in Fig. 2b. These materials will be discussed later in the paper.

After data acquisition, we split the whole data into the training-validation set (90%) and the test set (10%). The training-validation set is used for grid search with tenfold cross-validation to search for a proper tree depth of the random forests models. We find that a maximum depth of 12 layers (with 100 estimators) is reasonable for obtaining a good balance between bias and variance. A deeper tree would not improve the model performance. After deciding on the maximum tree depth, we also utilize the training-validation set to further refine the machine learning models. For an unbiased

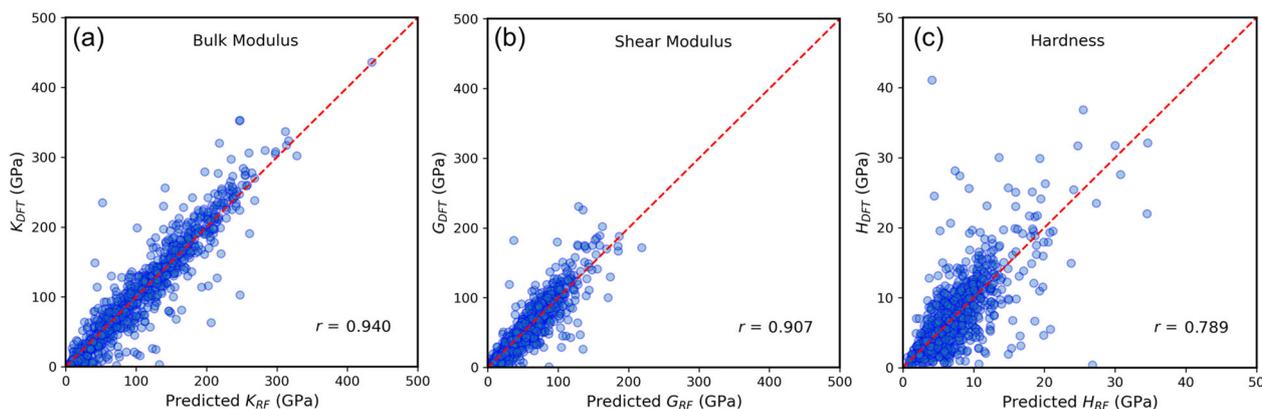


Fig. 3 Evaluation of the random forests (RF) models. The Pearson correlation coefficient (r) is utilized as a metric for **a** bulk modulus (K), **b** shear modulus (G), and **c** hardness (H). The machine learning models are trained to predict, respectively, K and G , and both can achieve $r > 0.9$ when applied to the test set based on density functional theory (DFT) calculations. H_{DFT} is computed using K_{DFT} and G_{DFT} with Tian's empirical formula. The predicted hardness H_{RF} is also obtained by Tian's formula with K_{RF} and G_{RF} as inputs, which results in an inferior correlation coefficient as anticipated.

evaluation of the final model performance, we use the test set and consider the metric of Pearson correlation coefficient r :

$$r = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^n (x_i - \bar{x})^2} \sqrt{\sum_{i=1}^n (y_i - \bar{y})^2}}. \quad (2)$$

Here x_i is the machine-learning predicted value for a single entry, and y_i is the corresponding "actual" value computed by DFT. \bar{x} and \bar{y} represent, respectively, the mean values of the predicted and the actual (DFT) values in the test set, which contains $n \sim 1000$ sample points. The r value can range between -1 and 1 , and $r = 1$ means that the prediction is 100% accurate.

Figure 3 shows the r value plots using the test set. For bulk and shear moduli (Fig. 3a, b), the data distribution follows closely the $r = 1$ dashed line. The r values of K and G are, respectively, 0.940 and 0.907 (and their coefficients of determination r^2 are, respectively, 0.885 and 0.822). For the hardness H_{RF} in Fig. 3c, we note that the prediction is not obtained directly from a machine learning model. Instead, we use the machine learning predicted K_{RF} and G_{RF} with Tian's empirical formula in Eq. (1) to compute H_{RF} . Therefore, the data distribution in Fig. 3c is more dispersing with an r value ~ 0.79 , which is slightly inferior as expected. We also note that a higher r value (or r^2 score) could be achieved by including additional features like volume, crystal symmetry, and cohesive energy, as done in previous studies^{32–37}. However, here we do not consider features that require additional measurements or calculations but focus only on features that can be derived directly from the chemical formula, in order to achieve efficient large-scale materials discovery.

Our machine learning models also provide information on feature importance to help reveal features that are more correlated with the bulk or shear moduli. Among the 60 features in our study, the atomic radius and d electron occupation are the most important ones. For example, the average atomic radius of a given compound and the bulk modulus exhibit a negative correlation with $r \sim -0.24$. Similarly, the r value between the largest atomic radius and the bulk modulus is $r \sim -0.41$. The results indicate that, in general, a smaller crystal unit cell will favor a higher bulk modulus. This is consistent with the facts that most superhard materials consist of light and small elements like Be, B, C, N, and O and that diamond has the smallest volume per unit cell among all crystalline materials. On the other hand, the d electron occupation rate is positively correlated with the bulk modulus, with a r value ~ 0.58 . This corresponds to the fact that many ultraincompressible materials are transition-metal borides like ReB_2 and Os_2B_3 ^{44,45}.

With the random forest models, we can predict quickly the mechanical properties of a given chemical formula. Here we apply our models to ternary B-C-N compounds by enumerating a series of $\text{B}_x\text{C}_y\text{N}_z$ compositions, with $x, y, z \in \{1, 2, 3, \dots\}$. Figure 4 shows the predicted triangular graphs, where the corner points correspond to unitary elemental compounds. For example, the pure boron phase in Fig. 4c is predicted to have $H_{\text{RF}} \sim 30$ GPa, which could be regarded as the hardness for α -B, β -B, γ -B, or tetragonal B_{52} . For the pure carbon phase, the predicted $H_{\text{RF}} \sim 90$ GPa could be related to cubic or hexagonal diamond (lonsdaleite). One caveat is that the relatively high hardness predicted near the pure nitrogen phase may be unrealistic; this is due to small bulk moduli of nitrogen-dominated compounds, which leads to a large Pugh's ratio k and a high hardness when Tian's model in Eq. (1) is used. If we implement more data selection rules by restricting the K and G values to be >50 GPa, then the artifact near the pure nitrogen phase could be avoided. However, this would cause overestimation in the overall mechanical properties.

Figure 4 also shows that B-C-N compositions with a 1:1 B:N ratio can result in several superhard compounds with hardness >60 GPa. For example, the predicted hardness values of BC_2N and BC_4N by machine learning are 74 and 65 GPa, respectively. These predictions are consistent with previous experimental findings of superhardness in BC_2N (76 or 62 GPa)^{18,19} and BC_4N (68 GPa)¹⁹, which are synthesized under high-pressure and high-temperature conditions.

First-principles calculation of structural and electronic properties

Motivated by the machine learning results in Fig. 4, we next employ evolutionary prediction with Universal Structure Predictor: Evolutionary Xtallography (USPEX) to search for potential superhard structures around the region with a B:N ratio $\sim 1:1$. The calculations are performed under an applied pressure of 15 GPa to help locate stable structures of smaller volumes and larger hardness. We first consider 15 trial chemical formulae with even number of valence electrons, including BC_3N , BC_5N , BC_6N , BC_3N_2 , $\text{B}_2\text{C}_3\text{N}$, $\text{B}_4\text{C}_5\text{N}_2$, etc., with a single-formula unit cell. However, most of the structures we found are graphite-like structures with sp^2 bonding, so they are not superhard. On the other hand, we find a diamond-like structure with sp^3 bonding for $\text{B}_2\text{C}_3\text{N}$ (Fig. 5b), which exhibits a hardness value >40 GPa and a relatively low formation energy as discussed later.

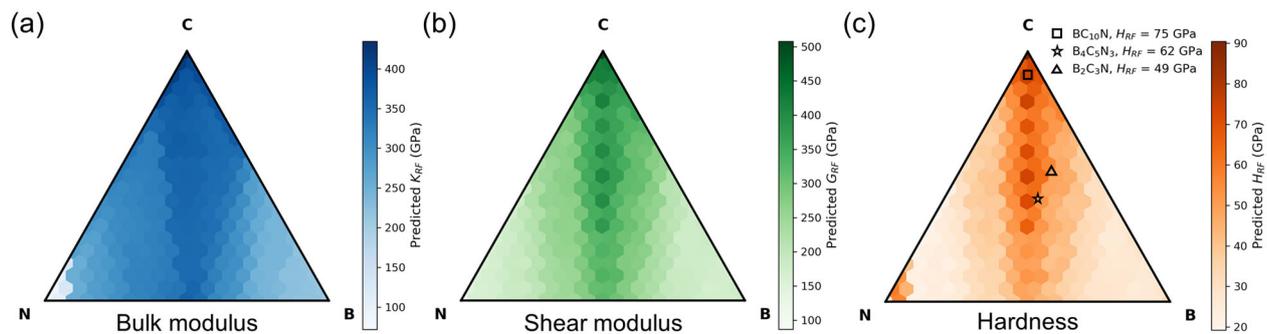


Fig. 4 Triangular graphs from machine learning prediction. B-C-N ternary graphs for **a** bulk modulus (K), **b** shear modulus (G), and **c** hardness (H), predicted by random forests (RF) machine learning models. **c** indicates that a 1:1 B-N composition ratio can lead to various superhard compounds such as BC_2N ($H_{RF} = 74$ GPa) and BC_4N ($H_{RF} = 65$ GPa). The hardness of three proposed superhard compounds, $BC_{10}N$, $B_4C_5N_3$, and B_2C_3N , are consistent with the DFT results. The triangular graphs are visualized by using the Python Ternary Plots library⁸⁷.

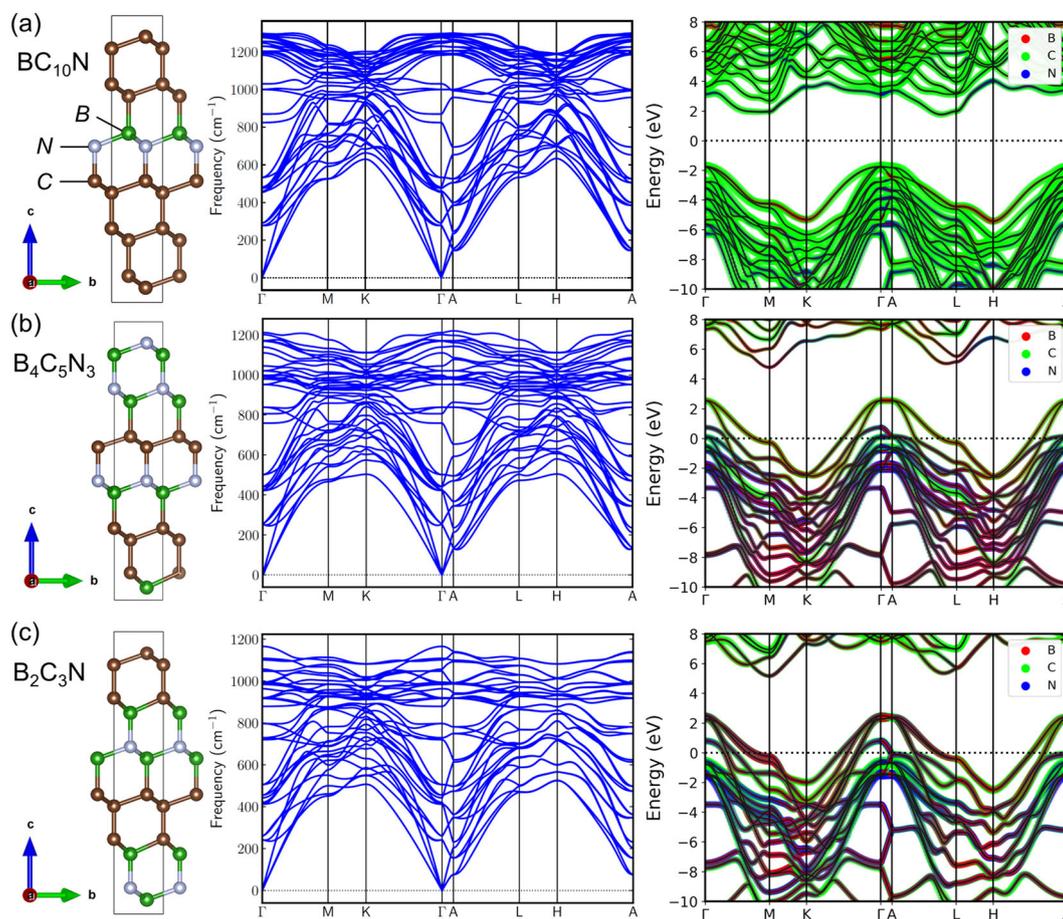


Fig. 5 Evolutionary algorithm and density functional theory calculations. Theoretical crystal structures (left panels), phonon dispersion spectra (middle panels), and electronic band structures (right panels) for **a** $BC_{10}N$, **b** $B_4C_5N_3$, and **c** B_2C_3N . $BC_{10}N$ is a wide band gap insulator, while $B_4C_5N_3$ and B_2C_3N are both metals. All three compounds are dynamically stable (i.e., without negative phonon modes). The crystal structures are visualized by the VESTA software⁸⁸.

B_2C_3N is hexagonal with a superlattice structure along the (111) direction of cubic diamond. By comparing the $1 \times 1 \times 2$ supercell of B_2C_3N (i.e., $B_4C_6N_2$), we find that such structure is similar to the $1 \times 1 \times 2$ BC_4N ⁴⁶ (i.e., $B_2C_8N_2$) with 2 carbon atoms replaced by 2 boron atoms, or the structure of $1 \times 1 \times 3$ BC_2N ⁴⁷ (i.e., $B_3C_6N_3$) with 1 nitrogen replaced by 1 boron. In fact, such atomic replacement is also the case of boron-substituted diamond BC_5 ⁴⁸ ($H \sim 70$ GPa) in a 12-atom unit cell, with 2 carbon atoms replaced by 2 boron

atoms. If one further replaces a carbon by boron in BC_5 , the resulting B_2C_4 (BC_2)⁴⁹ structure is also superhard ($H \sim 56$ GPa). Similarly, based on the 12-atom unit cell of diamond, the aforementioned superhard structure of BC_2N ⁴⁷ (BC_4N ⁴⁶) also can be generated by replacing 6 (4) carbons with 3 (2) BN pairs.

Using a 12-atom unit cell with a 1:1 B:N ratio, we first create the structure of $BC_{10}N$ (Fig. 5a) by replacing 2 carbon atoms in diamond by 1 pair of BN. Our random forests models predict that

Table 1. Structural and mechanical properties calculated by density functional theory (DFT).

Formula	$a(=b)$	c	ρ	K	G	E	k	ν	A^U	H_{RF}	H_{DFT}	ΔE
BC ₁₀ N	2.533	12.453	0.173	417	487	1052	1.166	0.080	0.058	75	87	79.5
B ₄ C ₅ N ₃	2.558	12.722	0.166	370	368	829	0.995	0.125	0.221	62	60	141.3
B ₂ C ₃ N	2.570	12.832	0.164	354	298	697	0.840	0.172	0.792	49	43	155.9
Diamond	2.527	12.379	0.175	432	518	1110	1.199	0.072	0.044	90	94	0
Cubic BN	2.564	12.559	0.168	373	383	856	1.025	0.118	0.172	50	64	0

The properties include lattice parameters $a(=b)$ and c (Å) in 12-atom hexagonal unit cell, density ρ (atom per Å³), bulk modulus K (GPa), shear modulus G (GPa), Young's modulus E (GPa), Pugh's ratio k , Poisson's ratio ν , universal elastic anisotropy A^U , hardness H (GPa), and formation energy ΔE (meV per atom). H_{DFT} is computed using K_{DFT} and G_{DFT} with Tian's empirical formula.

BC₁₀N has bulk and shear moduli equal to $K_{RF} = 379$ GPa and $G_{RF} = 422$ GPa, respectively, which corresponds to a hardness $H_{RF} = 75$ GPa by Tian's model. Using a similar rule, we also generate a superhard composition B₄C₅N₃, by replacing 1 carbon with 1 boron in B₃C₆N₃. B₄C₅N₃ is predicted to have $K_{RF} = 359$ GPa and $G_{RF} = 369$ GPa, with $H_{RF} = 62$ GPa. Other superhard B-C-N compounds also could be generated in a similar way. For example, B₃C₇N₂ and B₂C₉N could be obtained, respectively, from B₃C₆N₃ and B₂C₈N₂ by atomic substitution. Before we shift the focus to first-principles DFT validation of machine learning results, some comments are in order: (i) We have considered boron substitution in an 8-atom unit cell of BC₂N (i.e., B₂C₄N₂) to obtain B₃C₃N₂. However, we find that B-C-N compounds in an 8-atom unit cell have higher formation energies >300 meV per atom, which is consistent with early study on BC₂N by Chen et al.⁵⁰ (ii) We did not consider nitrogen substitution, because such structures tend to be thermodynamically more unstable.

We next discuss DFT calculations of three superhard B-C-N phases predicted by machine learning: BC₁₀N, B₄C₅N₃, and B₂C₃N. The structure of B₂C₃N (Fig. 5c) is discovered by USPEX using a 6-atom unit cell. For BC₁₀N (Fig. 5a) and B₄C₅N₃ (Fig. 5b), they are generated by using the aforementioned rule of atomic substitution in a 12-atom diamond unit cell. We have performed additional USPEX calculations for BC₁₀N and B₄C₅N₃ but did not find any lower enthalpy structure. The structures in Fig. 5 can be treated as superlattices along the (111) direction of cubic diamond. All structures are trigonal systems with the hexagonal lattice space group $P3m1$ (No. 156), which has 6 independent elastic constants C_{ij} : C_{11} , C_{12} , C_{13} , C_{14} , C_{33} , C_{44} , ($C_{66} = (C_{11} - C_{12})/2$). There are 4 necessary and sufficient mechanical stability conditions based on Born's criteria⁵¹: $C_{11} > C_{12}$, $C_{44} > 0$, $C_{13}^2 < \frac{1}{2}C_{33}(C_{11} + C_{12})$, and $C_{14}^2 < \frac{1}{2}C_{44}(C_{11} - C_{12})$. The three structures in Fig. 5 all fulfill these criteria, so they are mechanically stable.

Figure 5 middle panels show phonon spectra for the corresponding structures in the left panels, and all structures are dynamically stable without negative modes. Their phonon dispersions are fairly similar, due to the similarity in the crystal structures. Among them, BC₁₀N has the highest phonon frequency above 1250 cm⁻¹, while B₄C₅N₃ and B₂C₃N have slightly lower phonon frequencies near 1200 cm⁻¹ at the top of the phonon bands. We note that cubic diamond has the highest phonon frequency above 1300 cm⁻¹. The results indicate that phonon bands are softened with increasing B/N content. This phonon softening is consistent with the phenomenon observed in boron-incorporated diamond BC₅⁵².

The electronic band structures are shown accordingly in Fig. 5 right panels. BC₁₀N exhibits a wide band gap ~3.5 eV, so it is a superhard insulator. On the other hand, the electron-deficient B₄C₅N₃ and B₂C₃N are metals, where their valence band maximums are shifted toward the conduction bands. Due to the similarity of phonon and electronic dispersion relations between

B₄C₅N₃/B₂C₃N and BC₅^{48,52,53}, superconductivity may be observed in the predicted B₄C₅N₃ and B₂C₃N compounds as well. However, discussion related to superconducting properties is beyond the scope of this study.

The elastic constants C_{ij} computed by DFT for the structures in Fig. 5 can be utilized to derive other mechanical properties, such as the bulk modulus (K), shear modulus (G), and Young's modulus (E). The Vicker's hardness H also can be calculated by Eq. (1) with K and G as input. Details of the calculation results are given in Table 1. Notably, BC₁₀N has supreme mechanical properties with a hardness of 87 GPa. Compared to diamond, BC₁₀N is expected to show better performance in higher temperature and humidity environments.

Table 1 also indicates that, when the B/N content increases, the mechanical strengths like bulk and shear moduli as well as hardness will tend to decrease, which is consistent with the trend of phonon softening. The computed Cauchy pressures ($=C_{12} - C_{44}$)⁵⁴ of BC₁₀N, B₄C₅N₃, and B₂C₃N are all negative (and equal to -405, -209, and -121 GPa, respectively), which suggests their brittle properties and strong covalent bondings. In addition, the Pugh's ratio ($k = G/K$)⁴² is >0.571 for all three ternary compounds, indicative of their brittle properties as well. By increasing the B/N content, other B-C-N compounds like B₄C₅N₃ and B₂C₃N can become more ductile. Table 1 also demonstrates that the density ρ (atom per Å³) is positively correlated with mechanical strength, while the Poisson's ratio ν and universal elastic anisotropy A^{US} have negative correlations with elastic moduli and hardness.

Finally, we evaluate the thermodynamic stability of the proposed B_xC_yN_z compounds, by calculating the formation energy ΔE (also shown in Table 1):

$$\Delta E = \frac{E(\text{B}_x\text{C}_y\text{N}_z) - yE(\text{C}) - zE(\text{BN}) - (x - z)E(\text{B})}{x + y + z} \quad (3)$$

which is the difference between the total energy E of B_xC_yN_z and the atomically weighted reference total energies E of diamond, cubic BN, and α -B. The formation energies of the three compounds are positive, which is consistent with previous study on BC₂N⁵⁰, suggesting that ΔE increases due to B-C and C-N bondings. Among the three proposed compositions, BC₁₀N has the least number of B-C and B-N bonds, and it has the lowest formation energy <100 meV per atom. Since ΔE of BC₁₀N is smaller than those of BC₂N and BC₄N, it is likely that BC₁₀N can be synthesized without extreme conditions⁵⁶, using, e.g., low-temperature plasma methods.

In summary, we have built random forest models to predict bulk and shear moduli by using target elastic properties in the Materials Project database³⁸. The machine learning models utilize only materials features that can be derived directly from a given chemical formula, so they are suitable for large-scale materials characterization and discovery. We have applied the resulting models to B-C-N compounds to search for superhard ternary materials. The machine-learning predicted ternary graphs indicate

that a 1:1 B:N ratio can lead to various promising superhard materials with hardness >40 GPa. We also have utilized evolutionary structure prediction together with first-principles DFT calculations to further validate the machine learning results. We have proposed three potential superhard ternary compounds— BC_{10}N , $\text{B}_4\text{C}_5\text{N}_3$, and $\text{B}_2\text{C}_3\text{N}$ —and fully characterized their properties using first-principles calculations. In predicted ternary compounds, BC_{10}N is a wide band gap semiconductor, while $\text{B}_4\text{C}_5\text{N}_3$ and $\text{B}_2\text{C}_3\text{N}$ show metallic behavior. Among them, BC_{10}N has a hardness value ~ 87 GPa and a relatively low formation energy. Therefore, BC_{10}N may be synthesized without high-pressure and high-temperature conditions, for example, using low-temperature plasma methods. Once synthesized, BC_{10}N will have a wide range of potential applications in extreme environments.

METHODS

Machine learning model

Data acquisition. There exist several online computational materials databases, such as AFLOW⁵⁷, Materials Project³⁸, NOMAD Encyclopedia⁵⁸, and the Open Quantum Materials Database (OQMD)⁵⁹. Here we use the Materials Project³⁸, which provides open access to various computed properties of known and predicted crystalline compounds. The corresponding Python Materials Genomics (Pymatgen) library⁶⁰ is utilized to extract the target properties of bulk modulus (K) and shear modulus (G). We have neglected samples whose Voigt and Reuss modulus values differ by >50 GPa; this class of samples are typically quasi-two-dimensional materials, which are not the focus of our study. The Materials Project³⁸ database also contains high-pressure phases and artificial crystal structures, which can exhibit extreme values of bulk and shear moduli. Therefore, we exclude those extreme outliers and focus on 10,421 selected compounds with K and G values both in the ranges of 0–550 GPa. The chemical compositions and their target properties of bulk and shear moduli for the 10,421 compounds are written as a Python dictionary object saved in a JSON file.

Feature generation. To build a supervised learning model using only chemical composition as input, we generate features (or descriptors) based on a compound's chemical formula. By following ref. ⁶¹, we consider features related to stoichiometric attributes, elemental properties, orbital occupations, and ionic levels. Part of the features can be generated with the Python library MATMINER⁶². We do not consider structural or electronic features like crystal symmetry, volume, melting point, band gap, etc. While including these additional features could improve the model performance, these information is a priori unknown for new compounds. To expedite materials discovery, we thereby do not include features that require additional first-principles calculations.

First, the stoichiometric features are computed using the L^p norm $\|x\|_p = (\sum_i |x_i|^p)^{1/p}$, where x_i is element i 's atomic fraction. These attributes capture the changes in atomic fraction, independent of the actual elements. As an example, the $p=2$ norm of Fe_2O_3 is $\|x\|_2 = \left(\left(\frac{2}{5}\right)^2 + \left(\frac{3}{5}\right)^2\right)^{1/2} \approx 0.721$ ⁶¹. Here we consider 3 stoichiometric features, including the $p=0$ norm (i.e., the number of chemical components), and the $p=2,3$ norms. The $p=1$ norm is equal to unity regardless of the chemical composition, so it is not considered. In addition, we do not find an apparent model improvement with more higher-order norms ($p > 3$), so they are not included.

Second, the elemental features are computed using the minimum, maximum, and range for properties of each element present, as well as the values of fraction-weighted mean $\bar{f} = \sum_i x_i f_i$ and average deviation $\hat{f} = \sum_i |f_i - \bar{f}|$. Here f_i is the property of element i , and x_i is the atomic fraction. We consider the following 10 properties: atomic number, atomic mass, element column number, row number, atomic radius, electronegativity, and the numbers of valence electrons in s , p , d , and f orbitals, respectively. Therefore, there are 50 elemental property features (=5 values \times 10 properties). Using again Fe_2O_3 as an example⁶¹, for the "atomic number" property, $\bar{f} = \frac{2}{5}(26) + \frac{3}{5}(8) = 15.2$ and $\hat{f} = \frac{2}{5}|26 - 15.2| + \frac{3}{5}|8 - 15.2| = 8.64$.

Third, 4 orbital-occupation features are computed using the fraction-

weighted average of the number of valence electrons, respectively, in s , p , d , and f orbitals, divided by the fraction-weighted average of the total number of valence electrons. For example, Fe_2O_3 's p -orbital occupation feature is $F_p = \frac{2/5 \times (0+3/5 \times 4)}{2/5 \times (8)+3/5 \times (6)} \approx 0.353$ ⁶¹.

Finally, 3 features are based on ionic levels. The first is a Boolean number denoting whether it is possible to form a neutral ionic compound, by assuming that each element takes exactly one of its common charge states. The other two features are based on the "ionic character" of a chemical bond: $I(\chi_i, \chi_j) = 1 - \exp(-(\chi_i - \chi_j)^2/4)$, where χ_i and χ_j are electronegativities for elements i and j , respectively. In Pauling scale, fluorine has the highest electronegativity value of $\chi_{\text{F}} = 3.98$, and francium has the lowest electronegativity value of $\chi_{\text{Fr}} = 0.70$. The two features we consider are, respectively, the maximum ionic character I between any two elements in a compound, and the mean ionic character $\bar{I} = \sum_{i,j} x_i x_j I(\chi_i, \chi_j)$.

In total, 60 features are created. To simplify the training task, we do not consider additional feature engineering such as degree-2 polynomials, which otherwise could lead to thousands of features and cause overfitting.

Model training, validation, and application. For regression task, we choose the random forests algorithm^{63,64}, which is a tree-based ensemble method. A random forests model builds multiple decision trees, by taking a random sample with replacement from the training set and a random subset of features to split tree nodes. The results averaged over individual trees serve as the final predictions, which help reduce variance and improve accuracy. However, without restriction on the tree depth, the model can become very deep and cause overfitting. Therefore, we constrain the pruning parameter of tree depth to regularize the models.

The model training is implemented with the SCIKIT-LEARN library⁶⁵. We use 90% of our samples as the training and validation set, which is then used to determine the tree depth by tenfold cross-validation. The remaining 10% is the test set used for an unbiased evaluation of the final model. We build two separate models to predict the bulk modulus (K) and shear modulus (G), respectively. We do not train a model for predicting the Vickers hardness (H), as the target hardness value is not as widely available as K and G . On the other hand, there exist several empirical models for evaluating H ^{66–71}, based on physical properties such as bond length, bond strength, electronegativity, and covalent radius. Here we adopt hardness models that require only bulk and shear moduli as inputs^{43,72}, so that our regression results of K and G can be employed directly to predict H . Both empirical models lead to similar prediction results.

After training and evaluation, we apply the models to predict mechanical properties of B-C-N compounds and search for superhard ternary materials. For candidate compositions identified with superhardness (i.e., $H \geq 40$ GPa), we then perform crystal structure prediction (CSP) and first-principles calculations to further validate the machine learning predictions.

Crystal structure prediction

CSP concerns finding the stable structure of a compound knowing only its chemical formula^{73–75}. In principle, this is achieved by locating the minimum of the Gibbs free energy $G_{\text{free}} = U + PV - TS$, where U is the total energy, P is the pressure, V is the volume, T is the temperature, and S is the entropy. In practice, the entropy and temperature effects are often neglected, and only the enthalpy $H_{\text{free}} = U + PV$ is minimized. The minima of the potential energy surface correspond to different stable and metastable structures, which could be stabilized under different P - T conditions.

CSP requires an accurate estimation of the system's total energy U (usually from first-principles calculation) and an efficient optimization technique. Here we utilize the highly efficient implementation of evolutionary algorithm in USPEX^{76–78}. Evolutionary algorithm is a population-based optimization technique using biological evolution concepts, such as mutation, recombination, and selection. Candidate solutions are individuals in a population, which will evolve after applications of the above operators and selection by a fitness function. For a given composition, we examine over at least 1000 structures. The first generation of structures are randomly created. Subsequent generations are created with 20% from random structures and 80% from heredity, soft mutation, and transmutation operators.

First-principles calculation

Our first-principles DFT calculations are performed with the Vienna Ab initio Simulation Package (VASP)^{79,80}. The Monkhorst–Pack sampling scheme⁸¹ is used with a Γ -centered k -point mesh of $21 \times 21 \times 5$ (resolution = $0.02 \times 2\pi$ per \AA) points in the Brillouin zone. The convergence criteria of self-consistent and structural relaxation calculations are set to 10^{-6} eV per unit cell and 10^{-3} eV per \AA , respectively. We adopt a plane wave energy cutoff of 520 eV, which is sufficient to converge the DFT total energy difference $< 10^{-4}$ eV per atom. For each crystal structure, we first fully relax the lattice parameters and atomic positions. After structure relaxation, we then compute the corresponding mechanical, electronic, and phonon properties. All calculations use projector augmented wave^{82,83} pseudopotentials and the Perdew–Burke–Ernzerhof generalized gradient approximation functional⁸⁴.

We employ the strain–stress method⁸⁵ in VASP to compute the elastic constants C_{ij} , which in turn can determine the bulk and shear moduli using the Vogit–Reuss–Hill formula^{39–41}. Phonon dispersion spectra are computed using the PHONOPY package⁸⁶. Density functional perturbation theory with $2 \times 2 \times 1$ supercells is adopted to evaluate the second-order force constants.

DATA AVAILABILITY

The target data for training machine learning models are saved in a JSON file downloadable at https://github.com/weichihub/ML_B-C-N_data. Data for crystal structure prediction and density functional theory calculations are available upon request from the authors.

CODE AVAILABILITY

The Python codes needed for creating the feature data and for reproducing our machine learning results are available at https://github.com/weichihub/ML_B-C-N.

Received: 3 November 2020; Accepted: 17 June 2021;

Published online: 21 July 2021

REFERENCES

- Friedrich, A., Winkler, B., Juarez-Arellano, E. A. & Bayarjargal, L. Synthesis of binary transition metal nitrides, carbides and borides from the elements in the laser-heated diamond anvil cell and their structure–property relations. *Materials* **4**, 1648 (2011).
- Zhao, Z., Xu, B. & Tian, Y. Recent advances in superhard materials. *Annu. Rev. Mater. Res.* **46**, 383 (2016).
- Yeung, M. T., Mohammadi, R. & Kaner, R. B. Ultracompressible, superhard materials. *Annu. Rev. Mater. Res.* **46**, 465 (2016).
- Kwashnin, A. G., Allahyari, Z. & Oganov, A. R. Computational discovery of hard and superhard materials. *J. Appl. Phys.* **126**, 040901 (2019).
- Le Godec, Y., Courac, A. & Solozhenko, V. L. High-pressure synthesis of superhard and ultrahard materials. *J. Appl. Phys.* **126**, 151102 (2019).
- Haines, J., Léger, J. M. & Bocquillon, G. Synthesis and design of superhard materials. *Annu. Rev. Mater. Res.* **31**, 1 (2001).
- Zhao, Z., Xu, B. & Tian, Y. Recent advances in superhard materials. *Annu. Rev. Mater. Res.* **46**, 383 (2016).
- Kurakevych, O. O. Superhard phases of simple substances and binary compounds of the B–C–N–O system: from diamond to the latest results (a review). *Phys. Rev. B* **31**, 139 (2009).
- Zhang, W., Chong, Y. M., He, B., Bello, I. & Lee, S.-T. in *Comprehensive Hard Materials* (ed. Sarin, V. K.) 607–639 (Elsevier, 2014).
- Solozhenko, V. L., Kurakevych, O. O., Andrault, D., Le Godec, Y. & Mezouar, M. Ultimate metastable solubility of boron in diamond: synthesis of superhard diamondlike BC₅. *Phys. Rev. Lett.* **102**, 015506 (2009).
- Li, Q. Superhard and superconducting structures of BC₅. *J. Appl. Phys.* **108**, 023507 (2010).
- Baker, P. A. Computational predictions and microwave plasma synthesis of superhard boron-carbon materials. *Materials* **11**, 1279 (2018).
- Uemura, N., Shirai, K., Eckert, H. & Kunstmann, J. Structure, nonstoichiometry, and geometrical frustration of α -tetragonal boron. *Phys. Rev. B* **93**, 104101 (2016).
- Baker, P. A. First-principles predictions and synthesis of B₅₀C₂ by chemical vapor deposition. *Sci. Rep.* **10**, 4454 (2020).
- Chakrabarty, K. et al. Superhard boron-rich boron carbide with controlled degree of crystallinity. *Materials* **13**, 3622 (2020).
- Liu, X. Synthesis and characterization of new “BCN” diamond under high pressure and high temperature conditions. *Cryst. Growth Des.* **11**, 1006 (2011).
- Zhang, H., Yao, S. & Widom, M. Predicted phase diagram of boron-carbon-nitrogen. *Phys. Rev. B* **93**, 144107 (2016).
- Solozhenko, V. L., Dub, S. N. & Novikov, N. V. Mechanical properties of cubic BC₂N, a new superhard phase. *Diam. Relat. Mater.* **10**, 2228 (2001).
- Zhao, Y. Superhard B–C–N materials synthesized in nanostructured bulks. *J. Mater. Res.* **17**, 3139 (2002).
- Wang, S. Novel superhard B–C–O phases predicted from first principles. *Phys. Chem. Chem. Phys.* **18**, 1859 (2016).
- Liu, C., Chen, M., He, J., Yu, S. & Liang, T. Superhard B₂CO phases derived from carbon allotropes. *RSC Adv.* **7**, 52192 (2017).
- Li, Q., Wang, J., Zhang, M., Li, Q. & Ma, Y. Superhard-driven search of the covalent network in the B₃NO system. *RSC Adv.* **5**, 35882 (2015).
- Bhat, S. High-pressure synthesis of novel boron oxynitride B₆N₄O₃ with sphalerite type structure. *Chem. Mater.* **27**, 5907 (2015).
- Steele, B. A. & Oleynik, I. I. Ternary inorganic compounds containing carbon, nitrogen, and oxygen at high pressures. *Inorg. Chem.* **56**, 13321 (2017).
- Schmidt, J., Marques, M. R. G., Botti, S. & Marques, M. A. L. Recent advances and applications of machine learning in solid-state materials science. *Npj Comput. Mater.* **5**, 83 (2019).
- Zhou, T., Song, Z. & Sundmacher, K. Big data creates new opportunities for materials research: a review on methods and applications of machine learning for materials design. *Engineering* **5**, 1017 (2019).
- Himanen, L., Geurts, A., Foster, A. S. & Rinke, P. Data-driven materials science: status, challenges, and perspectives. *Adv. Sci.* **6**, 1900808 (2019).
- Chibani, S. & Couderc, F. Machine learning approaches for the prediction of materials properties. *APL Mater.* **8**, 080701 (2020).
- Saal, J. E., Oliyynyk, A. O. & Meredig, B. Machine learning in materials discovery: confirmed predictions and their underlying approaches. *Annu. Rev. Mater. Res.* **50**, 49 (2020).
- Cai, J., Chu, X., Xu, K., Li, H. & Wei, J. Machine learning-driven new material discovery. *Nanoscale Adv.* **2**, 3115 (2020).
- Meredig, B. Combinatorial screening for new materials in unconstrained composition space with machine learning. *Phys. Rev. B* **89**, 094104 (2014).
- Furmanchuk, A., Agrawal, A. & Choudhary, A. Predictive analytics for crystalline materials: bulk modulus. *RSC Adv.* **6**, 95246 (2016).
- de Jong, M. A statistical learning framework for materials science: application to elastic moduli of k-nary inorganic polycrystalline compounds. *Sci. Rep.* **6**, 34256 (2016).
- Isayev, O. et al. Universal fragment descriptors for predicting properties of inorganic crystals. *Nat. Commun.* **8**, 15679 (2017).
- Evans, J. D. & Couderc, F. Predicting the mechanical properties of zeolite frameworks by machine learning. *Chem. Mater.* **29**, 7833 (2017).
- Tehrani, A. M. Machine learning directed search for ultracompressible, superhard materials. *J. Am. Chem. Soc.* **140**, 9844 (2018).
- Avery, P. Predicting superhard materials via a machine learning informed evolutionary structure search. *Npj Comput. Mater.* **5**, 89 (2019).
- Jain, A. The Materials Project: a materials genome approach to accelerating materials innovation. *APL Mater.* **1**, 011002 (2013).
- Voigt, W. *Lehrbuch der Kristallphysik*, Vol. 1 (Teubner, 1928).
- Reuß, A. Berechnung der fließgrenze von mischkristallen auf grund der plastizitätsbedingung für einkristalle. *Z. Angew. Math. Mech.* **9**, 49 (1929).
- Hill, R. The elastic behaviour of a crystalline aggregate. *Proc. Phys. Soc. A* **65**, 349 (1952).
- Pugh, S. F. XCLL relations between the elastic moduli and the plastic properties of polycrystalline pure metals. *London Edinburgh Dublin Philos. Mag. J. Sci.* **45**, 823 (1954).
- Tian, Y., Xu, B. & Zhao, Z. Microscopic theory of hardness and design of novel superhard crystals. *Int. J. Refract. Hard Met.* **33**, 93 (2012).
- Burrage, K. C., Lin, C.-M., Chen, W.-C., Chen, C.-C. & Vohra, Y. K. Experimental and computational studies on superhard material rhenium diboride under ultrahigh pressures. *Materials* **13**, 1657 (2020).
- Burrage, K. C., Lin, C.-M., Chen, W.-C., Chen, C.-C. & Vohra, Y. K. Electronic structure and anisotropic compression of Os₂B₃ to 358 GPa. *J. Condens. Matter Phys.* **32**, 405703 (2020).
- Luo, X. Refined crystal structure and mechanical properties of superhard BC₄N crystal: first-principles calculations. *J. Phys. Chem. C* **112**, 9516 (2008).
- Liu, L. et al. Hexagonal BC₂N with remarkably high hardness. *J. Phys. Chem. C* **122**, 6801 (2018).
- Yao, Y., Tse, J. S. & Klug, E. D. Crystal and electronic structure of superhard BC₅: first-principles structural optimizations. *Phys. Rev. B* **80**, 094106 (2009).
- Xu, L. Prediction of a three-dimensional conductive superhard material: diamond-like BC₂. *J. Phys. Chem. C* **114**, 22688 (2010).

50. Chen, S., Gong, X. G. & Wei, S.-H. Superhard pseudocubic BC₂N superlattices. *Phys. Rev. Lett.* **98**, 015502 (2007).
51. Mouhat, F. & Coudert, F. Necessary and sufficient elastic stability conditions in various crystal systems. *Phys. Rev. B* **90**, 224104 (2014).
52. Baker, P. A. et al. Computational predictions and microwave plasma synthesis of superhard boron-carbon materials. *Materials* **11**, 1279 (2018).
53. Li, Q. Superhard and superconducting structures of BC₅. *J. Appl. Phys.* **108**, 023507 (2010).
54. Eberhart, M. E. & Jones, T. E. Cauchy pressure and the generalized bonding model for nonmagnetic bcc transition metals. *Phys. Rev. B* **86**, 134106 (2012).
55. Ranganathan, S. I. & Ostoja-Starzewski, M. Universal elastic anisotropy index. *Phys. Rev. Lett.* **101**, 055504 (2008).
56. Aykol, M., Dwaraknath, S. S., Sun, W. & Persson, K. A. Thermodynamic limit for synthesis of metastable inorganic materials. *Sci. Adv.* **4**, eaaq0148 (2018).
57. Curtarolo, S. AFLOW: an automatic framework for high-throughput materials discovery. *Comput. Mater. Sci.* **58**, 218 (2012).
58. Draxl, C. & Scheffler, M. The NOMAD laboratory: from data sharing to artificial intelligence. *J. Phys. Mater.* **2**, 036001 (2019).
59. Saal, J. E., Kirklin, S., Aykol, M., Meredig, B. & Wolverton, C. Materials design and discovery with high-throughput density functional theory: the open quantum materials database (OQMD). *JOM* **65**, 1501 (2013).
60. Ong, S. P. Python Materials Genomics (PYMATGEN): a robust, open-source python library for materials analysis. *Comput. Mater. Sci.* **68**, 314 (2013).
61. Ward, L., Agrawal, A., Choudhary, A. & Wolverton, C. A general-purpose machine learning framework for predicting properties of inorganic materials. *npj Comput. Mater.* **2**, 1 (2016).
62. Ward, L. Matminer: an open source toolkit for materials data mining. *Comput. Mater. Sci.* **152**, 60 (2018).
63. Ho, T. K. The random subspace method for constructing decision forests. *IEEE Trans. Pattern Anal. Mach. Intell.* **2**, 832 (1998).
64. Amit, Y. & Geman, D. Shape quantization and recognition with randomized trees. *Neural Comput.* **9**, 1545 (1997).
65. Pedregosa, F. Scikit-learn: machine learning in python. *J. Mach. Learn. Res.* **12**, 2825 (2011).
66. Gao, F. Hardness of covalent crystals. *Phys. Rev. Lett.* **91**, 015502 (2003).
67. Šimůnek, A., & Vackár, J. Hardness of covalent and ionic crystals: first-principle calculations. *Phys. Rev. Lett.* **96**, 085501 (2006).
68. Li, K., Wang, X., Zhang, F. & Xue, D. Electronegativity identification of novel superhard materials. *Phys. Rev. Lett.* **100**, 235504 (2008).
69. Lyakhov, A. O. & Oganov, A. R. Evolutionary search for superhard materials: methodology and applications to forms of carbon and TiO₂. *Phys. Rev. B* **84**, 092103 (2011).
70. Niu, H., Niu, S. & Oganov, A. R. Simple and accurate model of fracture toughness of solids. *J. Appl. Phys.* **125**, 065105 (2019).
71. Mazhnik, E. & Oganov, A. R. A model of hardness and fracture toughness of solids. *J. Appl. Phys.* **126**, 125109 (2019).
72. Chen, X.-Q., Niu, H., Li, D. & Li, Y. Modeling hardness of polycrystalline materials and bulk metallic glasses. *Intermetallics* **19**, 1275 (2011).
73. Wang, Y. & Ma, Y. Crystal structure prediction at high pressures. *J. Chem. Phys.* **140**, 040901 (2014).
74. Graser, J., Kauwe, S. K. & Sparks, T. D. Machine learning and energy minimization approaches for crystal structure predictions: a review and new horizons. *Chem. Mater.* **30**, 3601 (2018).
75. Oganov, A. R., Pickard, C. J., Zhu, Q. & Needs, R. J. Structure prediction drives materials discovery. *Nat. Rev. Mater.* **4**, 331 (2019).
76. Oganov, A. R. & Glass, C. W. Crystal structure prediction using ab initio evolutionary techniques: principles and applications. *J. Chem. Phys.* **124**, 244704 (2006).
77. Glass, C. W., Oganov, A. R. & Hansen, N. USPEX—evolutionary crystal structure prediction. *Comput. Phys. Commun.* **175**, 713 (2006).
78. Lyakhov, A. O., Oganov, A. R., Stokes, H. T. & Zhu, Q. New developments in evolutionary structure prediction algorithm USPEX. *Comput. Phys. Commun.* **184**, 1172 (2013).
79. Kresse, G. & Furthmüller, J. Efficiency of ab-initio total energy calculations for metals and semiconductors using a plane-wave basis set. *Comput. Mater. Sci.* **6**, 15 (1996).
80. Kresse, G. & Furthmüller, J. Efficient iterative schemes for ab initio total-energy calculations using a plane-wave basis set. *Phys. Rev. B* **54**, 11169 (1996).
81. Monkhorst, H. J. & Pack, J. D. Special points for Brillouin-zone integrations. *Phys. Rev. B* **13**, 5188 (1976).
82. Blöchl, P. E. Projector augmented-wave method. *Phys. Rev. B* **50**, 17953 (1994).
83. Kresse, G. & Joubert, D. From ultrasoft pseudopotentials to the projector augmented-wave method. *Phys. Rev. B* **59**, 1758 (1999).
84. Perdew, J. P., Burke, K. & Ernzerhof, M. Generalized gradient approximation made simple. *Phys. Rev. Lett.* **77**, 3865 (1996).
85. Page, Y. L. & Saxe, P. Symmetry-general least-squares extraction of elastic data for strained materials from ab initio calculations of stress. *Phys. Rev. B* **65**, 104104 (2002).
86. Togo, A. & Tanaka, I. First principles phonon calculations in materials science. *Scr. Mater.* **108**, 1 (2015).
87. Harper, M. et al. python-ternary: ternary plots in python. *Zenodo* <https://doi.org/10.5281/zenodo.594435>.
88. Momma, K. & Izumi, F. VESTA 3 for three-dimensional visualization of crystal, volumetric and morphology data. *J. Appl. Cryst.* **44**, 1272 (2011).

ACKNOWLEDGEMENTS

This research is supported by the U.S. National Science Foundation (NSF) under award OIA-1655280. The calculations were performed on the Frontera computing system at the Texas Advanced Computing Center. Frontera is made possible by NSF award OAC-1818253.

AUTHOR CONTRIBUTIONS

W.-C.C., J.N.S., D.Y., and C.-C.C. performed machine learning simulations and analyzed the results. W.-C.C. performed crystal structure prediction and density functional theory calculations. Y.K.V. and C.-C.C. conceived and supervised the project. W.-C.C. and C.-C.C. wrote the manuscript. All authors helped revise the manuscript.

COMPETING INTERESTS

The authors declare no competing interests.

ADDITIONAL INFORMATION

Correspondence and requests for materials should be addressed to C.-C.C.

Reprints and permission information is available at <http://www.nature.com/reprints>

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2021