Applied Physics A (2026) 132:744
https://doi.org/10.1007/s00339-026-09913-4

Applied Physics A
Materials Science & Processing

Load-dependent hardness prediction for materials using machine learning

Madhubanti Mukherjee[1,2] · Rampi Ramprasad[1] · Harikrishna Sahu[1]

Received: 18 May 2026 / Accepted: 5 July 2026

Abstract
Superhard materials are critical for wear-resistant and high-stress applications. Conventional approaches correlating hardness with elastic moduli derived from DFT calculations enable rapid screening but overlook the strong load dependence of hardness. In this work, machine learning (ML) models were developed using a large, curated dataset of load-dependent experimental Vickers hardness (H_v) measurements. Moderate correlation was observed between experimental and DFT-based H_v values, whereas a single-task ML model trained solely on experimental data outperformed multi-task models that combined experimental and computed data. The superior performance of the single-task model highlights that explicit inclusion of indentation load, along with compositional, electronic, and structural descriptors, is essential and sufficient for accurate hardness prediction, beyond what can be achieved using DFT-accessible bulk and shear moduli alone (or in tandem with experimental data). Specifically, the single-task Gaussian Process Regression (GPR) yields a train-test root mean squared error (RMSE) in a range of 2.73-2.90/3.16-3.91 GPa and R^2 in a range of 0.96-0.95/0.94-0.91 across five random splits of train and test sets. These results emphasize the importance of high-quality experimental data and explicit inclusion of measurement conditions, particularly load, in the development of reliable hardness prediction models.

Keywords Vickers hardness prediction · Load-dependent hardness · Superhard materials · Gaussian Process Regression (GPR) · Materials informatics

1 Introduction

Hardness is a key mechanical property that underpins technologies ranging from cutting tools and abrasives to protective coatings and structural components [1, 2]. Superhard materials, typically defined by Vickers hardness (H_v) above 40 GPa, [3–5] are essential for extreme environments where resistance to plastic deformation is critical. The experimental discovery of such materials is labor-intensive, which motivates the development of predictive models. Semiempirical approaches that estimate H_v from DFT-calculated elastic moduli enable high-throughput screening but rely on simplified assumptions that often fail under real conditions[6–9]. Machine learning (ML) models offer a flexible alternative with improved accuracy, yet most treat H_v as a static property [10–15], neglecting the influence of applied load.

While a few previous studies have incorporated load as an input parameter in machine learning models for hardness prediction and developed screening frameworks [16], the size and diversity of the dataset remain critical factors in assessing the applicability and reliability of such models. In order to ensure statistical robustness, it is important to utilize datasets that span a broad range of material compositions. Additionally, prediction-specific uncertainty quantification is particularly important for experimentally measured properties such as hardness, where variability arising from measurement conditions (such as load) can significantly influence reported values. In this context, probabilistic learning frameworks such as Gaussian Process Regression (GPR) provide reliable confidence estimation through principled uncertainty quantification. Another

✉ Harikrishna Sahu
harikrishnasahu89@gmail.com

[1] School of Materials Science and Engineering, Georgia Institute of Technology, Atlanta, GA 30332, USA

[2] Laboratory of Computational Chemistry and Biochemistry, Institute of Chemical Sciences and Engineering, Swiss Federal Institute of Technology (EPFL), Lausanne, Switzerland

Published online: 15 July 2026

important objective is to understand whether commonly used DFT-based hardness proxies meaningfully improve model performance or introduce bias when combined with experimental data. Furthermore, empirical hardness models derived from elastic moduli are themselves subject to inherent physical limitations. In this context, a multitask learning framework is a useful approach to assess the relative contribution of these proxies and to identify which, if any, offer closer agreement with experimentally measured hardness or add predictive value to the model. While hardness is a central metric for wear-resistant and structural applications, in advanced alloy designs such as Multi-Principal Element Alloys (MPEAs), it cannot be evaluated as an isolated property. Optimizing mechanical performance frequently requires balancing a well-documented trade-off between strength or hardness and ductility or elongation. Recent high-throughput and machine-learning frameworks have emphasized synergy-driven design strategies to simultaneously optimize these complementary properties. While the primary scope of this study focuses explicitly on capturing and predicting the experimentally measured, load-dependent Vickers hardness (H_v) response across diverse chemical spaces rather than multi-objective property optimization, acknowledging these co-dependent structural vectors remains critical for downstream material selection.

Here, we present a systematic evaluation of single-task versus multitask learning frameworks by explicitly including load as an input descriptor, allowing the model to capture its well-known dependence on hardness. Using a large experimental dataset spanning diverse crystal systems, compositions, and applied loads, we compared single-task (ST) and multi-task (MT) machine learning models built from purely experimental data and those incorporating DFT-based proxies to evaluate predictive performance. Notably, ST models trained solely on experimental data outperform all others, underscoring the limitations of DFT-based proxies in capturing hardness physics.

2 Methodology

2.1 Dataset curation and target space

The curated training database aggregates experimental Vickers hardness (H_v) values across wide compositional and structural domains. Crucially, to model the intrinsic physical phenomenon of the indentation size effect wherein the measured material hardness exhibits a non-linear dependence on the applied force, the exact experimental indentation load associated with each literature data point was integrated as an explicit descriptor alongside the chemical features.

To develop a reliable, load-dependent hardness prediction framework, a comprehensive dataset of experimental H_v measurements was curated, primarily from Refs. [16–18]. After removing duplicates with identical compositions and loads, the final dataset comprised 2,480 unique records covering 69 unique elements (the top 20 most frequent elements are shown in Fig. S1(a)) with compositions containing up to six elements (ternary: 1209, binary: 742, quaternary: 428, 5+ elements: 97 and unary: 4 as shown in Fig. S1(b)) and a total of 514 distinct chemical systems. The distribution of data across applied loads is depicted in Fig. 1(a). As shown, the measurements extend beyond 50 N, with 0-2 N (1,646 points), 2-4 N (364), 4-10 N (444), 10-50 N (24), and >50 N (2), representing a realistic range of indentation conditions for model training and validation.

2.2 Feature space

To map complex chemical information into a feature space, each material configuration is encoded using an array of 60 core compositional descriptors, generated using pymatgen [19]. This feature space adopts the mathematical formulations described in our previous study on superhard B-C-O systems [15]. For instance, descriptors such as the average number of valence electrons act as proxies for localized covalent bond densities and stability, and compositional or structural factors such as atomic-size mismatches and electronegativity differences capture lattice distortion effects and ionic bonding configurations, which directly govern resistance to plastic deformation. It is important to note that macroscopic hardness is strongly influenced by microstructural characteristics and processing parameters, such as fabrication route, grain size, porosity, and specific heat treatment profiles. Due to the heterogeneous nature of the reported experimental literature, standardized records of these processing pathways are largely absent from legacy datasets. Consequently, our descriptors focus primarily on robust compositional, electronic, and baseline structural properties. As highly curated, processing-aware material databases evolve, incorporating manufacturing-route descriptors will mark a vital step toward coupling intrinsic chemistry with extrinsic microstructural features in ML models.

2.3 Machine learning model architecture

Model development and training were performed using the Polymerize platform [20], and Gaussian Process Regression (GPR) was selected as the machine learning architecture. GPR is particularly well-suited for materials science datasets because it can efficiently capture highly complex, non-linear relationships in moderately sized tabular datasets without requiring the large volumes of data typically needed

by deep neural networks. Furthermore, GPR operates as a probabilistic learning framework, producing predictions as posterior distributions. This inherently provides a predictive standard deviation alongside each hardness prediction, enabling the direct extraction of mathematically rigorous 95% confidence intervals (prediction error bounds) for all datapoints.

3 Results

To establish a baseline for evaluating ML performance and integrate the computed data into MT learning, several widely used semiempirical models were examined that estimate H_v from DFT-calculated elastic moduli. Specifically, five models proposed by Tian [8], Chen [21], Jiang [22], and Teter [23] were considered:

$$H_v^1 = 0.92(G/B)^{1.137}G^{0.708} \quad (1)$$

$$H_v^2 = 2(G^3/B^2)^{0.585} - 3 \quad (2)$$

$$H_v^3 = 0.0963B \quad (3)$$

$$H_v^4 = 0.1475G \quad (4)$$

$$H_v^5 = 0.1769G - 2.899 \quad (5)$$

These five equations, referred to as Models 1–5 throughout the manuscript, relate H_v to the bulk modulus (B) and shear modulus (G), which represent resistance to volumetric and shear deformation, respectively. Using both in-house DFT calculations and data retrieved from the Materials Project [14, 24, 25], a dataset of 10,448 compositions was compiled, each assigned a nominal load of 0 N to reflect their theoretical nature.

Notably, 350 compositions overlap between the experimental and DFT-derived datasets, enabling direct comparison between measured and computed H_v values for model benchmarking. Scatter plots comparing experimental and predicted H_v obtained from the five semiempirical models, along with the corresponding Pearson correlation coefficients(r), are shown in Fig. 1(b). To ensure a meaningful comparison, the experimental data were restricted to measurements obtained under low indentation loads (typically $\leq$ 10 N), where plastic deformation more closely reflects the idealized conditions assumed in theoretical models [26–28]. The correlation between empirical predictions

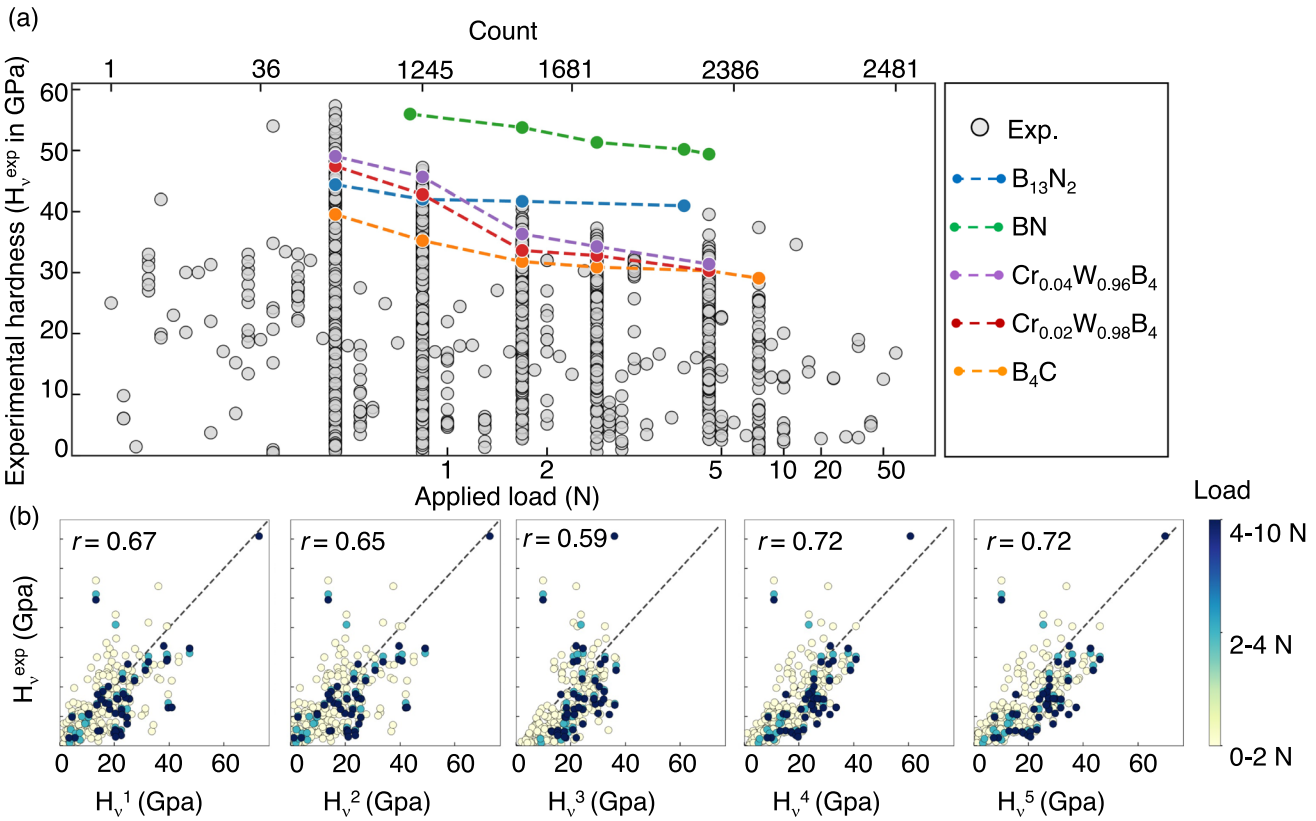

Fig. 1 Hardness data showing (**a**) the distribution of experimental hardness values with respect to applied load and (**b**) the correlation between experimental and theoretically calculated hardness obtained using different approaches

and experimental H_v ranged from 0.59 to 0.72 across the models. This moderate agreement highlights two fundamental limitations: (i) the static formulation of these models, which neglects the strong load dependence of hardness, and (ii) their inherent simplifying assumptions of isotropy, defect-free structures, and the absence of microstructural effects. To further examine load effects, we analyzed the variation of experimental H_v with applied load for representative materials (Fig. 1(a)). The results reveal that H_v can change by more than a factor of two with increasing load-behavior not captured by any empirical model, underscoring the need for predictive frameworks that explicitly learn and account for load-dependent effects.

To address the limitations of empirical approaches, a series of ML models was developed to establish a standardized modeling framework and evaluate the effect of dataset expansion and diversification on predictive performance. In total, six models were constructed: a ST model trained exclusively on experimental data, and five MT models that combine experimental data with computed hardness values obtained from the five semiempirical formulations given in (1)–(5).

For feature generation, 60 compositional, electronic, and structural descriptors were computed based on the elemental composition of each compound. These descriptors include atomic number, mass, electronegativity, atomic radius, *s*- and *p*-orbital electron counts, and periodic table coordinates. Statistical measures such as mean, range, and weighted averages were applied to encode compositional complexity into the model space. In addition to these descriptors, the indentation load was explicitly incorporated as an input feature in all models to capture its critical influence on H_v. By embedding indentation load directly into the descriptor space, the model will be able to capture the transition from ideal elastic response at low or zero loads to deformation at higher loads, regimes that are generally averaged in elastic-modulus based empirical approaches. All ML models were developed using Gaussian Process Regression (GPR) training algorithm implemented in PolymRize™ platform [20], an advanced framework for rapid model development and evaluation. Each model was trained with 5-fold cross-validation (CV) and multiple random train–test splits to ensure statistical robustness, and performance was assessed using the root mean square error (RMSE) and coefficient of determination (R^2).

Figure 2 summarizes the performance of the developed GPR model using only the experimental dataset. The learning curves in Fig. 2(a) show the average RMSE for the CV-train, CV-test, and independent test sets as a function of training set size. As expected, the RMSE values for both CV-test and test sets decrease with increasing training data, indicating improved generalization with larger datasets. Figure 2(b) presents the average RMSE from five different random splits, with error bars representing the standard deviation, while Fig. S2 shows the corresponding R^2. The mean RMSE values for the CV-test and test sets are 4.03 and 3.49 MPa, respectively, with corresponding R^2 values of 0.91 and 0.93. A representative parity plot for the train and test sets is shown in Fig. 2(c), and parity plots for all five random splits are provided in Fig. S2.

The performance of the five MT GPR models, which combine experimental data with computed hardness values obtained from the five semiempirical approaches, is shown in Fig. S3. As observed, the MT models incorporating computed data from the Chen and Tian formulations performed worse than the ST model, while the remaining three exhibited comparable RMSEs but lower R^2 values. These results indicate that incorporating empirically calculated H_v values does not improve predictive performance for the materials considered. The absence of load dependence in the semiempirical H_v values, along with their reliance on bulk and shear moduli, likely limits their physical relevance, yielding little to no improvement in MT learning performance. Several earlier ML models trained on hardness values derived from empirical correlations with DFT-calculated elastic moduli have exhibited consistently poorer performance relative to the ST model presented here [10, 11, 13–15]. To show how our input parameters relate to each other and to the material hardness, we performed a simple Pearson correlation analysis. Figure S4 (a) shows the top 10 features that have the highest correlation to hardness. Red bars show features that increase hardness when their values go up, while blue bars show features that have an inverse effect on hardness. Figure S4 (b) displays the cross-correlation among these 10 features, with some blocks showing strong correlation. Typically, an atom's weight, atomic number, and size are all physically related by the laws of the periodic table. We clarify that we purposely chose to keep all 60 compositional features and the load to develop the model to ensure inclusivity of all subtle but vital physical properties the model needs to learn to be robust. While high correlation can sometimes make it tricky for the model to rank individual feature importance, it does not affect the model's actual prediction accuracy or its ability to predict the hardness of entirely new, unseen materials. We note that the model is not expected to work well with extrapolation to loads or material chemistries outside the training domain, and the predictive performance may vary for materials with limited load coverage in the dataset or those exhibiting atypical or nonlinear deformation behavior.

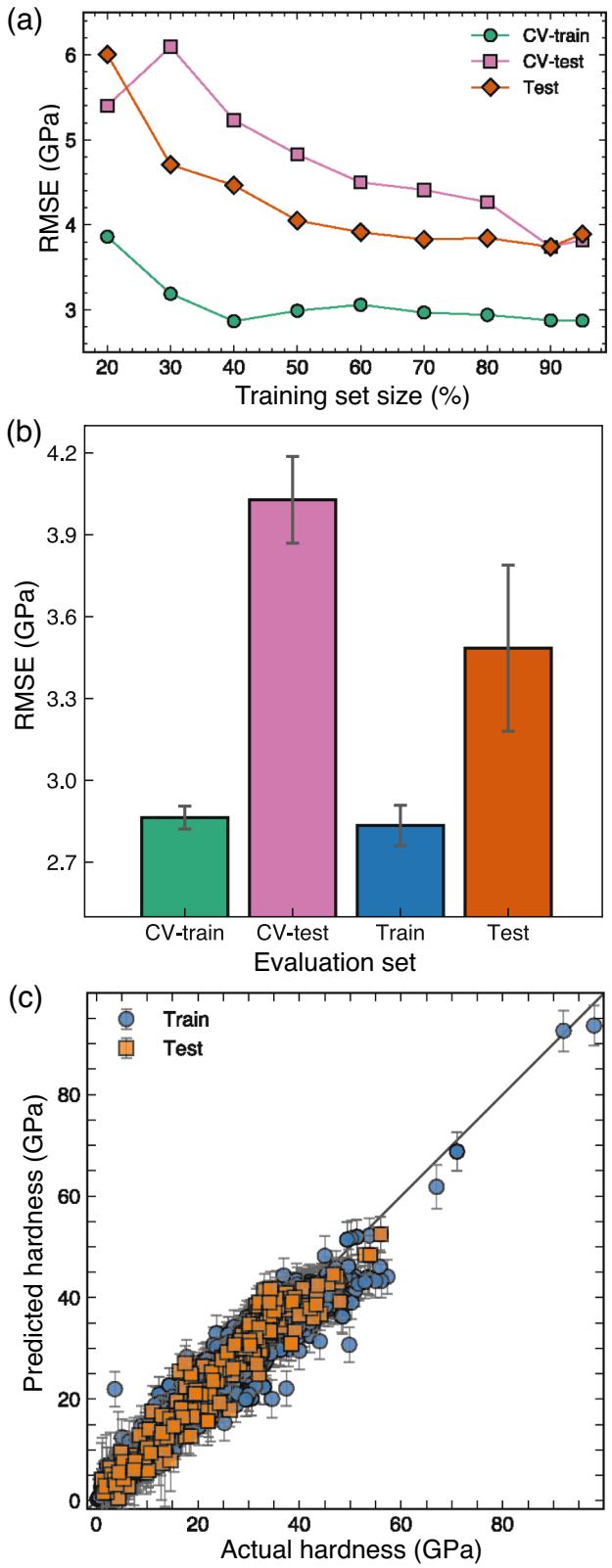

Fig. 2 Performance of the single-task GPR model. (**a**) Learning curve showing RMSE (GPa) for cross-validation (CV)-train, CV-test, and independent test sets with increasing training-set size. (**b**) Average RMSEs with standard deviations over five random splits for CV-train, CV-test, train, and independent test sets. (**c**) Representative parity plot comparing predicted and actual hardness values for one split

4 Conclusion

In this study, single-task and multi-task GPR models were developed to predict Vickers hardness (H_v) across a wide range of materials and indentation loads. The single-task model trained exclusively on experimental data achieved excellent predictive performance, with RMSE and R^2 values of 3.49 MPa and 0.93, respectively. In contrast, multi-task models that incorporated computed H_v values from five semiempirical approaches alongside the experimental data showed no improvement and, in some cases, performed worse. These findings demonstrate that elastic moduli alone are insufficient for accurate hardness prediction, highlighting the importance of high-quality, load-dependent experimental data, along with compositional, electronic, and structural descriptors, for developing reliable and physically meaningful hardness prediction models.

Supplementary information The Supplementary Material provides detailed performance analyses of the single-task GPR model trained solely on experimental data and the multitask models that integrate experimental and theoretical hardness data. The experimental hardness dataset, computational elastic moduli dataset, and hardness values calculated using five semiempirical models are available on GitHub at https://github.com/Ramprasad-Group/polyVERSE.

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1007/s00339-026-09913-4.

Acknowledgements The authors are grateful for the financial support of this work by the Office of Naval Research (grant number N00014-21-1-2258). The authors also thank ACCESS for computational support through allocation DMR080044.

Author Contributions M.M. was the primary architect of the project, including data collection, development of the machine learning model, and manuscript preparation. H.S. assisted in developing the ML model, provided guidance, and supervised the work. R.R. conceived the project and provided overall guidance.

Data Availability The experimental hardness dataset, computational elastic moduli dataset, and hardness values calculated using five semiempirical models are available on GitHub at https://github.com/Ramprasad-Group/polyVERSE.

Declarations

Conflicts of Interest There are no conflicts to disclose.

Ethical Approval Not applicable.

Consent for publication Not applicable.

Competing interests The authors declare no competing interests.

References

1. R.B. Kaner, J.J. Gilman, S.H. Tolbert, Designing superhard materials. Science **308**(5726), 1268–1269 (2005)
2. F. Gao, X. Qin, L. Wang, Y. He, G. Sun, L. Hou, W. Wang, Prediction of new superhard boron-rich compounds. J. Phys. Chem. B **109**(31), 14892–14895 (2005)
3. Z. Zhao, B. Xu, Y. Tian, Recent advances in superhard materials. Annu. Rev. Mater. Res. **46**(1), 383–406 (2016)
4. M.T. Yeung, R. Mohammadi, R.B. Kaner, Ultraincompressible, superhard materials. Annu. Rev. Mater. Res. **46**, 465–485 (2016)
5. A.G. Kvashnin, Z. Allahyari, A.R. Oganov, Computational discovery of hard and superhard materials. J. Appl. Phys. **126**(4), 040901 (2019)
6. X. Jiang, J. Zhao, X. Jiang, Correlation between hardness and elastic moduli of the covalent crystals. Comput. Mater. Sci. **50**(7), 2287–2290 (2011)
7. A. Šimůnek, J. Vackář, Hardness of covalent and ionic crystals: first-principle calculations. Phys. Rev. Lett. **96**, 085501 (2006)
8. F. Gao, J. He, E. Wu, S. Liu, D. Yu, D. Li, S. Zhang, Y. Tian, Hardness of covalent crystals. Phys. Rev. Lett. **91**, 015502 (2003)
9. A.M. Tehrani, J. Brgoch, Hard and superhard materials: a computational perspective. J. Solid State Chem. **271**, 47–58 (2019)
10. M. De Jong, W. Chen, R. Notestine, K. Persson, G. Ceder, A. Jain, M. Asta, A. Gamst, A statistical learning framework for materials science: application to elastic moduli of k-nary inorganic polycrystalline compounds. Sci. Rep. **6**(1), 34256 (2016)
11. O. Isayev, C. Oses, C. Toher, E. Gossett, S. Curtarolo, A. Tropsha, Universal fragment descriptors for predicting properties of inorganic crystals. Nat. Commun. **8**(1), 15679 (2017)
12. A. Mansouri Tehrani, A.O. Oliynyk, M. Parry, Z. Rizvi, S. Couper, F. Lin, L. Miyagi, T.D. Sparks, J. Brgoch, Machine learning directed search for ultraincompressible, superhard materials. J. Am. Chem. Soc. **140**(31), 9844–9853 (2018)
13. P. Avery, X. Wang, C. Oses, E. Gossett, D.M. Proserpio, C. Toher, S. Curtarolo, E. Zurek, Predicting superhard materials via a machine learning informed evolutionary structure search. NPJ Comput. Mater. **5**(1), 89 (2019)
14. W.-C. Chen, J.N. Schmidt, D. Yan, Y.K. Vohra, C.-C. Chen, Machine learning and evolutionary prediction of superhard bcn compounds. NPJ Comput. Mater. **7**(1), 114 (2021)
15. M. Mukherjee, H. Sahu, M.D. Losego, W.R. Gutekunst, R. Ramprasad, Informatics-driven design of superhard b–c–o compounds. ACS Appl. Mater. Interfaces. **16**, 10372–10379 (2024)
16. Z. Zhang, A. Mansouri Tehrani, A.O. Oliynyk, B. Day, J. Brgoch, Finding the next superhard material through ensemble learning. Adv. Mater. **33**, 2005112 (2021)
17. J.C. Hickey, J. Brgoch, The limits of proxy-guided superhard materials screening. Chem. Mater. **34**, 10003–10010 (2022)
18. R. Jaafreh, Y.S. Kang, J.-G. Kim, K. Hamad, Machine learning guided discovery of super-hard high entropy ceramics. Mater. Lett. **306**, 130899 (2022)
19. S.P. Ong, W.D. Richards, A. Jain, G. Hautier, M. Kocher, S. Cholia, D. Gunter, V.L. Chevrier, K.A. Persson, G. Ceder, Python materials genomics (pymatgen): A robust, open-source python library for materials analysis. Comput. Mater. Sci. **68**, 314–319 (2013)
20. Inc. Matmerize, PolymRize™: An informatics platform for polymer property prediction and design using machine learning (2025). https://polymrize.matmerize.com/
21. X.-Q. Chen, H. Niu, D. Li, Y. Li, Modeling hardness of polycrystalline materials and bulk metallic glasses. Intermetallics **19**, 1275–1281 (2011)
22. X. Jiang, J. Zhao, X. Jiang, Correlation between hardness and elastic moduli of the covalent crystals. Comput. Mater. Sci. **50**(7), 2287–2290 (2011)
23. D.M. Teter, Computational alchemy: The search for new superhard materials. MRS Bull. **23**, 22–27 (1998)
24. A. Jain, S.P. Ong, G. Hautier, W. Chen, W.D. Richards, S. Dacek, S. Cholia, D. Gunter, D. Skinner, G. Ceder, K.A. Persson, Commentary: The materials project: A materials genome approach to accelerating materials innovation. APL Mater. **1**, 011002 (2013)
25. M. De Jong, W. Chen, T. Angsten, A. Jain, R. Notestine, A. Gamst, M. Sluiter, C. Krishna Ande, S. Van Der Zwaag, J.J. Plata et al., Charting the complete elastic properties of inorganic crystalline compounds. Sci. Data **2**(1), 150009 (2015)
26. E. Williams, Hooke's law and the concept of the elastic limit. Ann. Sci. **12**(1), 74–83 (1956)
27. H. Lan, T. Venkatesh, On the relationships between hardness and the elastic and plastic properties of isotropic power-law hardening materials. Philos. Mag. **94**(1), 35–55 (2014)
28. V. Gorban, M. Lugovyi, D. Verbylo, Determination of material hardness characteristics at the elastic limit by instrumented indentation. Strength Mater. **57**(3), 567–572 (2025)