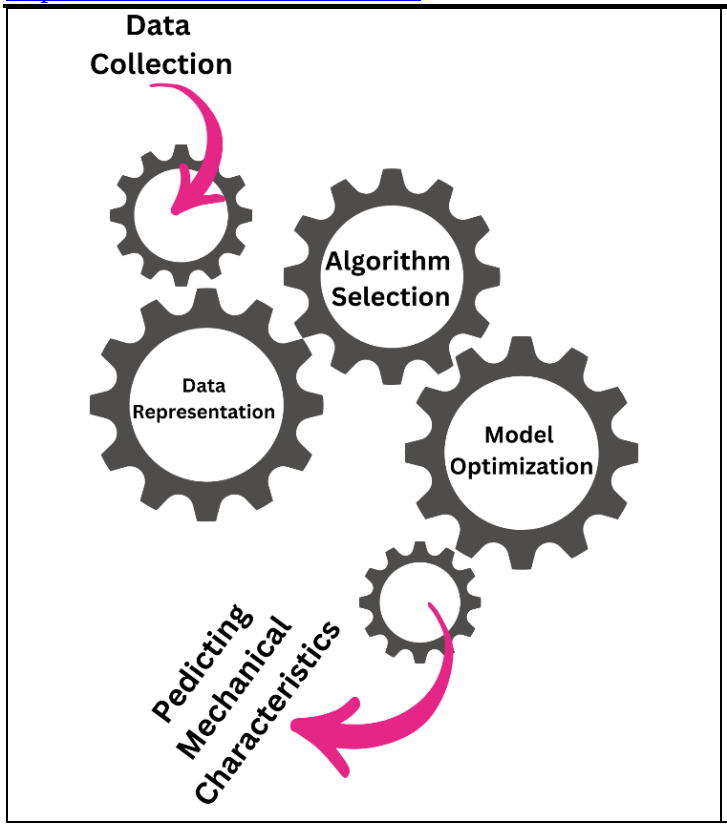
MOL2NET'23, Conference on Molecular, Biomedical & Computational Sciences and Engineering, 9th ed.

Innovation in Materials: Key Steps for Algorithm Selection in Predicting Mechanical Characteristics through Machine Learning

Galo Cerda-Mejía[a,b], Karel Diéguez-Santana[a], Juan M. Ruso[b]

[a] Universidad Regional Amazónica Ikiam, Parroquia Muyuna km 7 vía Alto Tena, 150150, Tena- Napo, Ecuador
[b] Soft Matter and Molecular Biophysics Group, Department of Applied Physics, University of Santiago de Compostela, 15782 Santiago de Compostela, Spain

Graphical Abstract	**Abstract**



The central importance of materials in society and their relationship with various properties is highlighted. The growing relevance of artificial intelligence (AI), especially machine learning (ML) and deep learning algorithms, in mechanical engineering and materials science is emphasized. The ability of AI to predict features and create innovative materials is highlighted. Furthermore, the crucial steps for applying ML in materials innovation are described, from data collection and cleaning to algorithm selection and optimization, emphasizing the importance of understanding the nature of data and model validation. Finally, a comprehensive overview of the integration of AI and ML in materials research is provided, highlighting their fundamental role in the optimization and prediction of mechanical properties.

Introduction

Materials play an extremely significant role in our lives, as they constitute the fundamental elements in creating essential tools for the advancement of our society. Multiple successful approaches have been developed to discover novel materials, aided by the recognition of the close relationship between the internal composition of materials and their various properties. These properties can be preliminarily grouped into categories such as mechanical, thermal, optical, electrical, chemical, nuclear, among others [1,4,5,6].

Artificial intelligence (AI), particularly machine learning (ML) and deep learning algorithms, is gaining significant relevance in the fields of mechanical engineering and materials science. This is due to its ability to anticipate material characteristics, devise innovative materials, and unveil novel mechanisms that transcend the realm of intuitive perceptions, where a working framework must be generated [1-3]. Machine learning models, trained using datasets containing information about materials, have the capability to provide rapid and accurate predictions regarding desired mechanical properties or behaviors. They can also discover compositions or structures that surpass the data used in training, within the design space [1,2].

The rapid evolution of computers in modern society has propelled their integration into various fields, progressively replacing human tasks. The growth in computing power and storage has made their use attractive in complex tasks. Artificial intelligence (AI), including machine learning (ML) and deep learning (DL), stands out, being employed in genomics, drug discovery, automation, and finance. ML, a prominent branch of AI, enables machines to learn without explicit programming, through supervised, unsupervised, and reinforcement learning. Feature engineering is crucial for optimizing ML models. Various algorithms, such as decision trees, SVM, and neural networks, have been developed. Advances in ML have led to autonomous vehicles, voice and face recognition, and breakthroughs in genomics [2, 9-12].

Methods

Selecting the right algorithm is essential for accurately predicting mechanical characteristics through machine learning in materials innovation. Understanding the nature of the data, comparing algorithms, validating with independent datasets, and adjusting parameters are key steps to optimize predictive capability. Therefore, the key steps are:

Steps	Description
Data Collection	Material data often exhibits deficiencies, repetitions, and inconsistencies due to experimental limitations. Data cleaning is essential to correct errors, using averages or statistical values to fill in missing information and merging duplicates. Specific programs verify the coherence of the data based on the range of values and relationships between variables, eliminating out-of-range or conflicting data. After cleaning, the data is ready for representation.
Data Representation	Data representation involves transforming information into formats more suitable for an algorithm. Although the data is usually numerical, its suitability for the algorithm is crucial. Similar to how we use equations or graphs to understand mathematical problems, machine learning algorithms need proper presentation of data for optimal learning. The quality of model performance improves with proper representation of the data.
Algorithm Selection	Artificial intelligence and machine learning are classified into supervised learning (which involves classification and regression with labeled data) and unsupervised (which includes clustering of unlabeled data). In materials research, reinforcement and active learning techniques are also applied. Notable algorithms include K-Nearest Neighbors (KNN), symbolic decision tree regression, and artificial neural networks. ✓ KNN ranks based on the majority of votes among the k closest points using distance measures such as Euclidean or Manhattan. ✓ Decision tree acts as a classifier that performs tests on attributes to make decisions, where the nodes represent tests and the branches represent attribute values. ✓ Symbolic regression is based on evolving candidate functions using chromosomes and nodes with operations and variables to fit experimental data. ✓ Artificial neural networks: model brain activity with nodes connected by directed links, with weights that spread activation and adjust to minimize errors. In addition, there are more advanced algorithms, such as random forests, kernel methods, convolutional neural networks, and generative adversarial networks. Current research focuses on automatic ML to simplify the implementation of these algorithms.
Model Optimization	The model optimization process involves evaluating a test model to select the best one. Three main sources of error, model bias, model variability, and irreducible errors, must be considered. Bias comes from incorrect assumptions in the algorithm, variability refers to sensitivity to small fluctuations in the training set. Poor performance indicates high bias (underfitting) or high variability (overfitting). Bias occurs when the model is not flexible enough or the data is insufficient. Variability occurs when the model becomes too complex. The key test is the successful application of the model to unseen data. A test set is used during training and validation, evaluating the effectiveness of the model. Cross-validation is reliable when samples are representative, but careful selection of methods is required to evaluate the transferability and applicability of the model in cases with diverse data sets.

Conclusion

The fusion between artificial intelligence and materials science has ushered in a revolutionary era of discovery and advancement. The predictive and design capabilities of machine learning models offer a systematic approach to materials research and development, marking a transformative shift in process efficiency and precision. As computing power evolves, the presence of artificial intelligence is consolidated in various disciplines, from genomics to finance. The steps outlined for applying machine learning in materials innovation underscore the importance of data quality and representation, algorithm selection, and model optimization. Looking ahead, exploring advanced algorithms and simplifying

implementation through automation promises to further expand the impact of artificial intelligence in materials science, paving the way for limitless innovation.

References

[1] Guo, K., Yang, Z., Yu, C. H., & Buehler, M. J. (2021). Artificial intelligence and machine learning in design of mechanical materials. Materials Horizons, 8(4), 1153-1172.
[2] Huang, J. S., Liew, J. X., Ademiloye, A. S., & Liew, K. M. (2021). Artificial intelligence in materials modeling and design. Archives of Computational Methods in Engineering, 28, 3399-3413.
[3] Sha, W., Guo, Y., Yuan, Q., Tang, S., Zhang, X., Lu, S., ... & Cheng, S. (2020). Artificial intelligence to power the future of materials science and engineering. Advanced Intelligent Systems, 2(4), 1900143.
[4] Butler, K.T., Davies, D.W., Cartwright, H. et al. Machine learning for molecular and materials science. Nature 559, 547–555 (2018). https://doi.org/10.1038/s41586-018-0337-2
[5] Mueller, Tim, Aaron Gilad Kusne, and Rampi Ramprasad. "Machine learning in materials science: Recent progress and emerging applications." Reviews in computational chemistry 29 (2016): 186-273.
[6] Gubernatis, J. E., & Lookman, T. (2018). Machine learning in materials design and discovery: Examples from the present and suggestions for the future. Physical Review Materials, 2(12), 120301.
[7] Liu, Y., Zhao, T., Ju, W., & Shi, S. (2017). Materials discovery and design using machine learning. Journal of Materiomics, 3(3), 159-177.
[8] Ramakrishna, S., Zhang, T. Y., Lu, W. C., Qian, Q., Low, J. S. C., Yune, J. H. R., ... & Kalidindi, S. R. (2019). Materials informatics. Journal of Intelligent Manufacturing, 30, 2307-2326.
[9] Agrawal, A., & Choudhary, A. (2019). Deep materials informatics: Applications of deep learning in materials science. Mrs Communications, 9(3), 779-792.
[10] Takahashi, K., & Tanaka, Y. (2016). Materials informatics: a journey towards material design and synthesis. Dalton Transactions, 45(26), 10497-10499.
[11] Bhadeshia, H. K. D. H. (2009). Neural networks and information in materials science. Statistical Analysis and Data Mining: The ASA Data Science Journal, 1(5), 296-305.
[12] Scott, D. J., Coveney, P. V., Kilner, J. A., Rossiny, J. C. H., & Alford, N. M. N. (2007). Prediction of the functional properties of ceramic materials from composition using artificial neural networks. Journal of the European Ceramic Society, 27(16), 4425-4435.
[13] Basheer, I. A., & Hajmeer, M. (2000). Artificial neural networks: fundamentals, computing, design, and application. Journal of microbiological methods, 43(1), 3-31.