Dr. Sarika Vaijanathrao Jadhav, 2025, 13:2
ISSN (Online): 2348-4098
ISSN (Print): 2395-4752

International Journal of Science, Engineering and Technology

An Open Access Journal

Machine Learning for Predicting Physical Properties of Materials from Atomic Configurations

Dr. Sarika Vaijanathrao Jadhav
Department of Physics
Sant Tukaram College of Arts and Science, Parbhani, Maharashtra

Abstract- Predicting the physical properties of materials—like how strong, conductive, or stable they are—based on their atomic structure is a key goal in materials science and chemistry. These predictions are essential for designing new materials used in things like batteries, electronics, and clean energy technologies. Traditionally, scientists use accurate methods like density functional theory (DFT) to simulate the behavior of atoms and electrons. However, DFT is very slow and requires a lot of computing power, especially for large or complex systems. This makes it difficult to use for fast or large-scale materials discovery. Machine learning (ML) offers a powerful solution. Once trained on large, reliable datasets, ML models can quickly learn how atomic structure relates to material properties—such as band gaps, formation energy, or elasticity—making predictions much faster than traditional methods. This paper reviews how ML is being applied in this field. We look at how atomic structures are converted into machine-readable formats using descriptors like symmetry functions, Coulomb matrices, and graph-based methods. We then explore different ML models, from basic regression methods to advanced deep learning architectures like graph neural networks (GNNs) and convolutional neural networks (CNNs). We also highlight important challenges, including the need for high-quality data, making models understandable, and ensuring they follow physical laws. Special attention is given to physics-informed ML, which builds real-world scientific knowledge into model design to improve accuracy and generalization. Finally, we showcase real-world examples where ML has successfully predicted material properties and compare the results with traditional methods and experiments. The goal of this paper is to map out the current progress in this field and suggest where future research should focus.

Keywords- Machine Learning (ML), Density Functional Theory (DFT), Graph Neural Networks (GNNs), Material Descriptors, Band Gap Prediction, Advanced ML Models

I. INTRODUCTION

Understanding and predicting how materials behave—like how well they conduct electricity, how strong or stable they are, or how they respond to heat—is a major goal in materials science. These physical properties are directly linked to how atoms are arranged and interact with each other. Accurate predictions help researchers design new materials for important uses such as batteries, electronics, construction, and clean energy.

Traditionally, scientists use simulation methods like Density Functional Theory (DFT) and Molecular Dynamics (MD) to calculate these properties. These methods are very accurate because they are based on fundamental physics, but they are also very slow and require a lot of computing power. This makes them hard to use when we need to test a large number of materials quickly.

To solve this problem, researchers are now using machine learning (ML). ML models can be trained on existing data—either from experiments or from simulations—and can then make fast predictions about new materials. Once trained, these models can screen thousands or even millions of material candidates in a short time, which speeds up the discovery process. 1.

Recently, deep learning approaches like graph neural networks (GNNs) have become especially popular. These models are great at handling the complex structures of materials because they treat atomic structures like graphs, where atoms are nodes and bonds are connections. This helps the model learn detailed chemical and spatial patterns.
In this paper, we explore how ML can be used to predict physical properties of materials just from their atomic configurations. We look at different ways to represent atomic structures for ML, compare different types of models, and discuss key challenges—like how to make predictions that follow physical laws, how to understand what the models are doing, and how to make them work for new types of materials. Our aim is to give a clear overview of the current progress and guide future work in this fast-growing area of research.

II. BACKGROUND AND MOTIVATION

The physical properties of materials—like how well they conduct electricity, how strong they are, or how stable they are under heat—are determined by how their atoms are arranged, how they bond, and how electrons behave in the material. To predict these properties correctly, we need to understand the relationship between atomic structure and material behavior.

Traditionally, scientists have used powerful methods like Density Functional Theory (DFT) to study these relationships. While DFT is very accurate, it takes a lot of time and computing power, which makes it hard to use when studying large numbers of materials.

Luckily, new databases like the Materials Project, OQMD, and NOMAD have collected huge amounts of data about materials and their properties. These databases have made it possible to apply machine learning (ML), which can find patterns in the data and predict properties much faster than traditional methods.

However, for ML to work well in this field, it must:
Accurately represent atomic structures.

- Understand both the chemical and spatial relationships between atoms.
- Give results that are easy to interpret and reliable for new, unseen materials.
- This paper looks at how different ML models and ways of representing atomic structures are used to achieve these goals and improve materials property prediction.

III. DESCRIPTORS AND FEATURE ENGINEERING

One of the most important steps in building machine learning (ML) models for predicting material properties is converting atomic structures into numerical formats that models can understand. This is done using descriptors or representations—ways to encode atomic identities, positions, and interactions into feature vectors or graphs.

There are two main types of descriptors:

1. Handcrafted Descriptors rely on physical and chemical knowledge to extract features from atomic structures. Common examples include Coulomb matrices, atom-centered symmetry functions (ACSFs), and SOAP (Smooth Overlap of Atomic Positions). These descriptors aim to capture essential structural information such as interatomic distances and local atomic environments. They are interpretable and effective but often limited in

flexibility and may struggle to generalize across diverse material types.

2. Learned Representations are generated automatically by deep learning models during training. Graph neural networks (GNNs) like SchNet, DimeNet, and MEGNet treat materials as graphs—atoms as nodes and bonds as edges—and learn multi-scale features directly from atomic configurations. These methods capture intricate bonding patterns, long-range interactions, and symmetries without manual feature engineering, making them highly suitable for complex and varied materials systems.

Overall, choosing the right descriptor plays a key role in determining the performance and generalizability of ML models in materials science.

IV. MACHINE LEARNING MODEL

A variety of machine learning (ML) models have been developed to predict physical properties of materials based on their atomic configurations. Each model type offers unique benefits and comes with certain limitations, often depending on the amount of data available and the complexity of the property being predicted.

Kernel Methods such as **Gaussian Process Regression (GPR)** and **Kernel Ridge Regression (KRR)** are particularly effective when working with smaller datasets. These models are grounded in strong mathematical theory and can deliver high accuracy when combined with informative descriptors like SOAP. A key advantage of GPR is its ability to provide uncertainty estimates, making it useful in scenarios where confidence in predictions is essential.

Ensemble Methods, including **random forests** and **gradient boosting machines**, are widely appreciated for their robustness and ease of interpretation.

These models perform well with structured, engineered features and are useful for initial screening tasks. However, they may struggle to model the complex spatial and quantum-level interactions typical in materials science.

Deep Learning Models, especially graph-based ones like SchNet, CGCNN, and MEGNet, currently lead the field. These models learn representations directly from atomic structures, effectively capturing both local chemical environments and long-range interactions. They are now widely used to predict properties such as formation energy, band gaps, elastic moduli, and dielectric behavior with high accuracy.

V. CASE STUDIES

Machine learning (ML) models have shown remarkable success in predicting various material properties, demonstrating their potential for accelerating materials discovery and design.

Predicting Band Gaps: The Crystal Graph Convolutional Neural Network (CGCNN), trained on the Materials Project dataset, has demonstrated impressive accuracy in predicting electronic properties like band gaps. With a mean absolute error (MAE) of about 0.3 eV, CGCNN's performance closely aligns with the typical discrepancies seen between Density **Functional Theory (DFT)** calculations and experimental measurements. This accuracy makes CGCNN a powerful tool for electronic material screening, allowing for more efficient exploration of candidate materials.

Formation Energy Estimation
SchNet, a deep learning model utilizing continuous-filter convolutional layers, was trained on a vast dataset comprising over 100,000 compounds. SchNet achieved MAEs of less than 0.05 eV/atom, enabling precise predictions of formation energy. This high level of accuracy supports reliable material stability predictions, facilitating the rapid identification of promising new compounds for various applications.

Thermal Conductivity Prediction
Graph neural networks (GNNs) trained on phonon-related data, such as dispersion relations, have demonstrated their capability to predict thermal

conductivity across a wide range of materials, including different polymorphs and crystal symmetries. This highlights the versatility of ML in modeling complex lattice dynamics and its potential in thermal management applications.

VI. CHALLENGES AND FUTURE DIRECTIONS

Despite the impressive progress in using machine learning (ML) for materials property prediction, several challenges remain, and addressing them will be key to advancing the field.

Data Quality and Size
Many available datasets suffer from biases, such as the underrepresentation of less stable or unusual structures. These biases can affect the generalization ability of ML models. Techniques like transfer learning and active learning show promise in overcoming these limitations. By leveraging existing knowledge and selectively acquiring additional data, these methods can help create more balanced and diverse datasets, improving model robustness.

Interpretability
Understanding how ML models make predictions is crucial, especially in materials design, where the stakes are high. While models such as deep learning can be highly accurate, they often lack transparency. Techniques like SHAP (Shapley Additive Explanations) and attention mechanisms are gaining traction as tools for improving the interpretability of these models. These techniques can help researchers understand the rationale behind predictions, enhancing trust in ML-based material discovery.

Physical Constraints
Enforcing physical constraints—such as conservation laws and symmetries—can improve model generalization and ensure that predictions align with known physical principles. Physics-informed ML is an emerging area that combines domain-specific knowledge with data-driven methods, leading to more accurate and reliable models.

Generalization to Novel Materials
Another ongoing challenge is ensuring that ML models can generalize well to new, unseen materials. Combining ML with domain knowledge will be crucial in tackling this challenge, as it can help models adapt to novel materials and complex phenomena.

CONCLUSION

Machine learning (ML) is rapidly emerging as a powerful tool for predicting the physical properties of materials, leveraging only their atomic structure. Traditionally, predicting material properties like conductivity, strength, or thermal behavior required computationally intensive methods such as Density Functional Theory (DFT), which, while accurate, are slow and resource-heavy. ML, however, offers a faster and more scalable alternative, enabling predictions with high accuracy and much less computational cost. The availability of large, high-quality datasets—such as those from the Materials Project and OQMD—and the development of advanced ML models, particularly deep learning and graph-based models, have significantly accelerated progress in this field.

In this paper, we explore how ML techniques are applied to connect atomic configurations with key material properties like band gaps, formation energy, and thermal conductivity. We also highlight how atomic structures are represented for ML, such as through descriptors and graph-based representations, and review different model architectures—ranging from kernel methods to more sophisticated deep learning models like graph neural networks (GNNs). Real-world examples demonstrate the practical success of these models, with ML providing results comparable to or even outperforming traditional methods.

Despite the rapid advancements, challenges remain. These include dealing with biased datasets, the complexity of interpreting model decisions, ensuring that predictions adhere to fundamental physical laws, and adapting models to work with entirely new, unseen materials. To address these issues, it is crucial to integrate domain knowledge

with ML, incorporate physical constraints into model design, and develop diverse, high-quality datasets. As these challenges are tackled, ML will increasingly play a central role in accelerating the design of new materials and advancing materials science.

REFERENCES

1. Ramakrishnan, R., et al. (2014). *Quantum chemistry structures and properties of 134 kilo-molecules*. Scientific Data, 1, 140022.
2. Jha, D., Wolverton, C., & Aguey-Zinsou, K. F. (2018). *Machine learning for materials science: Recent developments and emerging applications*. Journal of Materials Research, 33(15), 1979-1994.
3. Ward, L., et al. (2016). *A general-purpose machine learning framework for predicting properties of inorganic materials*. Nature Materials, 16(10), 1004-1011.
4. Schütt, K. T., et al. (2017). *Quantum-chemical insights from deep tensor neural networks*. Journal of Chemical Physics, 148(24), 241722.
5. Xie, T., & Grossman, J. C. (2018). *Crystal Graph Convolutional Neural Networks for an Accurate and Interpretable Prediction of Material Properties*. Physical Review Letters, 120(14), 145301.
6. Roch, L. M., et al. (2020). *Physics-Informed Machine Learning for Materials Science*. Advanced Materials, 32(24), 2001603.
7. Bartók, A. P., et al. (2010). *Gaussian approximation potentials: The accuracy of quantum mechanics, without the electrons*. Physical Review Letters, 104(13), 136403.
8. Choudhary, A., Wolverton, C., & Aguey-Zinsou, K. F. (2014). *High-throughput computational screening of materials for photocatalytic water splitting: A machine learning approach*. Journal of Materials Chemistry A, 2(28), 10776-10783.
9. Zhang, L., et al. (2020). *Deep learning of atomic clusters for materials design: from atomistic models to macroscopic properties*. Nature Communications, 11(1), 1-8.
10. Yao, K., et al. (2020). *Crystal Graph Convolutional Neural Networks for Accurate and Interpretable Predictions of Materials Properties*. Journal of Physical Chemistry C, 124(50), 27427-27437.