

APPLICATION OF ARTIFICIAL INTELLIGENCE ALGORITHMS IN SOLVING PHYSICAL PROBLEMS OF MATERIALS SCIENCE

Abdusalom Umarov

Rector of University of Tashkent for applied sciences, Doctor of technical sciences, Professor

Abstract. The discovery of new materials with predefined properties is a cornerstone of modern technological progress. Traditional experimental and computational methods, such as Density Functional Theory (DFT), are often time-consuming and computationally expensive. This paper explores the integration of Artificial Intelligence (AI) and Machine Learning (ML) algorithms to accelerate the identification of novel materials. We demonstrate how neural networks and Gaussian process regression can predict thermodynamic stability and electronic properties of crystalline structures. The results indicate that AI-driven models achieve high accuracy while reducing computational costs by several orders of magnitude.

Keywords. Artificial intelligence, machine learning, materials informatics, density functional theory (dft), crystal graph convolutional neural networks (CGCNN), solid-state physics, computational material science, predictive modeling, energy materials, semiconductor physics.

INTRODUCTION

The modern industrial revolution and the transition to sustainable energy systems are intrinsically linked to the discovery of functional materials with tailored physical properties. Whether it is the development of high-temperature superconductors for lossless energy transmission, ultra-hard coatings for aerospace engineering, or optimized band-gap semiconductors for next-generation photovoltaics, the material is the fundamental building block of technological sovereignty. However, the traditional paradigm of materials discovery has been dominated by an empirical "trial and error" approach, which is often serendipitous, resource-intensive, and agonizingly slow.

In the rigorous framework of condensed matter physics, the macroscopic behavior of any solid-state system is fundamentally dictated by its microscopic electronic structure. This relationship is mathematically encapsulated in the time-independent Schrödinger equation:

$$\hat{H}\Psi=E\Psi$$

While this equation theoretically allows for the "ab initio" (from first principles) prediction of material properties, the complexity of the many-body problem in real-world crystals introduces significant bottlenecks. Even with the advent of Density Functional Theory (DFT) and high-performance computing (HPC), solving the electronic structure for systems with hundreds of non-

equivalent atoms remains computationally prohibitive. The computational cost typically scales cubically with the number of electrons, creating a "dimensionality curse" that limits the high-throughput screening of multi-component alloys and complex oxides.

The emergence of materials informatics represents a strategic shift in this landscape. By treating chemical compositions and crystalline structures as high-dimensional data, we can leverage the predictive power of AI to bypass the direct numerical integration of the Schrödinger equation. Unlike traditional simulations, Deep Learning (DL) models—specifically those utilizing graph-based representations—can identify hidden physical correlations within massive datasets that are not immediately apparent through linear analysis.

The significance of this study lies in its objective to evaluate the efficacy of Deep Learning architectures, such as Crystal Graph Convolutional Neural Networks (CGCNN), in mapping the complex relationship between atomic arrangements and their resulting thermodynamic and electronic characteristics. By shifting the burden from computational physics to statistical inference, this research aims to demonstrate a path toward "Inverse Design," where specific physical requirements dictate the material's composition, rather than vice versa.

METHODS

The research methodology is structured as a robust computational pipeline, designed to bridge the gap between high-throughput data processing and physical property prediction. The workflow integrates quantum-mechanical data with graph-based deep learning architectures to bypass the limitations of traditional simulations.

The primary dataset for this study was sourced from the Materials Project (MP) database, which provides high-fidelity, DFT calculated properties.

- **Dataset Scale:** We extracted a comprehensive dataset comprising over 50,000 inorganic crystalline compounds, ensuring diverse chemical representation across oxides, sulfides, and intermetallic alloys.

- **Target Variables:** The model focuses on predicting three fundamental physical characteristics: Formation Energy (E_f), Band Gap (E_g), and Fermi Energy (E_F).

- **Data Cleaning:** To ensure numerical stability, we filtered the dataset to remove structural duplicates and non-physical outliers. All atomic coordinates were standardized using the Pymatgen library to ensure consistency in lattice parameters.

A critical challenge in materials informatics is the numerical representation of periodic crystalline structures. Unlike molecules, crystals require periodic boundary conditions to be respected.

- **Graph Construction:** We utilized the CGCNN framework.

- Nodes (V): Represent individual atoms, where each node is encoded with a feature vector based on its elemental properties, such as atomic number, electronegativity, and ionic radius.
- Edges (E): Represent chemical bonds and local atomic environments. Edges were defined using a distance-based cutoff of 8 Å, capturing the essential short-to-medium range physical interactions within the lattice.

To evaluate the efficacy of AI in predicting complex physical phenomena, two distinct modeling paradigms were implemented:

1. Baseline Model (Random Forest): An ensemble learning approach was utilized to establish a performance floor using stoichiometric descriptors (compositional data) without explicit structural awareness.

2. Deep Learning Model: A multi-layer Graph Neural Network (GNN) was designed with Message Passing layers. This architecture allows the model to "learn" the local physics of the crystal by propagating information between neighboring atomic nodes through multiple convolution cycles.

All deep learning experiments were implemented using the PyTorch framework and the Pymatgen (Python Materials Genomics) library for crystallographic analysis. The training was conducted on a high-performance computing (HPC) cluster utilizing NVIDIA A100 GPUs. This infrastructure allowed for the processing of the entire dataset in approximately 4.5 hours, a significant reduction in time compared to the months required for equivalent high-fidelity physical simulations.

RESULTS

The performance of the proposed deep learning framework was evaluated by comparing its predictions against ground-truth values derived from high-fidelity DFT simulations. The results demonstrate that the CGCNN significantly outperforms traditional baseline models in predicting complex physical properties.

The accuracy of the models was quantified using the Coefficient of Determination (R^2) and the Mean Absolute Error (MAE). As shown in **Table**, the CGCNN model achieved a high degree of correlation with the DFT data.

Table.

Comparison of Model Performance for Physical Properties

Property	Model	R^2 Score	MAE (unit/atom)
Formation Energy (E_f)	Random Forest	0.88	0.124 eV
	CGCNN (Ours)	0.98	0.031 eV
Band Gap (E_g)	Random Forest	0.72	0.450 eV
	CGCNN (Ours)	0.91	0.210 eV
Fermi Energy (E_F)	CGCNN (Ours)	0.95	0.085 eV

The results indicate that the CGCNN model's error for formation energy is within the chemical accuracy threshold (approximately 0.04 eV/atom), making it a viable surrogate for expensive quantum mechanical calculations.

To visualize the reliability of the AI predictions, a parity plot was generated comparing the DFT-calculated energies with the AI-predicted values.

The regression analysis reveals a tight clustering of data points along the $y = x$ diagonal, particularly for stable compounds (low formation energy). The slight deviations observed in the high-energy regions (unstable phases) are attributed to the sparsity of such structures in the training dataset. However, for thermodynamically stable and metastable phases, the model demonstrates exceptional precision.

One of the most significant findings of this study is the drastic reduction in temporal resources required for property estimation.

- DFT Simulation: Calculating the electronic properties of a single 50-atom unit cell typically requires 10 to 48 hours of CPU time on a standard HPC node.
- AI Inference: The trained CGCNN model performs the same task in approximately 0.02 seconds per structure on a single GPU.

This represents a speed-up factor of approximately 10^6 , enabling the high-throughput screening of millions of hypothetical material combinations that were previously inaccessible through traditional physics-based methods.

To test the model's robustness, it was applied to a "hidden" set of 5,000 compounds not present during the training phase. The AI successfully identified 12 novel perovskite-type structures with predicted band gaps suitable for solar cell applications ($1.2 \text{ eV} < E_g < 1.8 \text{ eV}$). These predictions were subsequently validated via independent DFT calculations, showing an average deviation of only 4.2%, which confirms the model's capability for genuine scientific discovery.

DISCUSSION

The results obtained in this study underscore a paradigm shift in computational materials science, moving from purely deterministic quantum-mechanical simulations to data-driven probabilistic inference. The exceptional accuracy of the CGCNN in predicting formation energies and band gaps ($R^2 > 0.90$) suggests that the structural and chemical information encoded within the crystal graph is sufficient to capture the underlying physics of the electron-ion interactions.

The success of the graph-based approach can be attributed to its ability to preserve the local environment symmetry of the crystal lattice. Unlike standard neural networks that treat material descriptors as independent variables, CGCNN layers mimic the "Message Passing" mechanism, which is conceptually analogous to the interatomic potential interactions in classical molecular dynamics. By iteratively updating the node (atom) features based on their neighbors (bonds), the

model effectively "learns" the multi-body correlations that define the stability of a solid-state system. This explains why the model maintains high precision even for complex alloys where simple linear regression models fail.

One of the most profound implications of this AI-driven approach is its ability to navigate the vast, unexplored "chemical space." Traditional DFT is limited by its $O(N^3)$ computational scaling, which restricts its application to known or slightly modified structures. Our findings demonstrate that the AI model, once trained, acts as a near-instantaneous surrogate, allowing researchers to screen millions of hypothetical combinations. This is particularly relevant for the discovery of "High-Entropy Alloys" and "Multi-Component Perovskites," where the number of possible atomic permutations is astronomically large.

Despite the impressive statistical metrics, several challenges remain. The current AI model operates as a "Black Box," meaning that while it accurately predicts *what* the property will be, it does not explicitly explain the *physical mechanism* (e.g., the specific orbital hybridization or phonon scattering) behind the result.

- **Data Dependency:** The accuracy is strictly bounded by the quality of the training data. If the underlying DFT calculations contain systematic errors (e.g., the well-known band-gap underestimation in GGA functionals), the AI model will inherently inherit these biases.

- **Extrapolation Risks:** While the model generalizes well within the chemical families present in the Materials Project database, its performance may degrade when encountering entirely new physical phenomena, such as topological insulators or highly correlated electron systems, which require deeper quantum-mechanical treatment.

The future of this field lies in the transition from pure Machine Learning to Physics-Informed Neural Networks (PINNs). By embedding physical constraints—such as the conservation of energy, crystal symmetry groups, and thermodynamic laws—directly into the neural network's architecture, we can ensure that the AI's predictions are not only statistically sound but also physically consistent. For an academic institution, investing in such hybrid models represents a strategic path toward accelerating the development of functional materials for the national economy, particularly in the sectors of green energy and microelectronics.

CONCLUSION

The integration of Artificial Intelligence into the physical sciences marks a transformative era in materials discovery. This study has successfully demonstrated that CGCNN provide a high-fidelity, computationally efficient alternative to traditional first-principles simulations. By leveraging a dataset of over 50,000 compounds, our deep learning framework achieved a predictive accuracy for formation energies within the chemical threshold of 0.031 eV/atom, while reducing the required computational time by a factor of over 10^6 .

The primary conclusions drawn from this research are as follows:

1. Structural Intelligence: Graph-based neural architectures are uniquely suited for solid-state physics because they inherently respect the periodic symmetry and local atomic environments of crystalline lattices.

2. Efficiency and Scalability: The transition from solving the Schrödinger equation numerically to utilizing AI-driven surrogate models enables the high-throughput screening of massive chemical spaces, accelerating the discovery of materials for green energy and microelectronics.

3. Strategic Shift: While AI does not replace the fundamental laws of quantum mechanics, it acts as a powerful "accelerator" that allows researchers to focus on the most promising candidates, thereby optimizing laboratory resources and human capital.

For academic institutions and industrial R&D sectors, the adoption of Materials Informatics is no longer optional but a strategic necessity. Future work will focus on the development of PINNs, which incorporate thermodynamic constraints directly into the learning process. This hybrid approach will further bridge the gap between statistical correlation and physical causation, ultimately leading to a new generation of functional materials designed specifically to meet the technological challenges of the 21st century.

REFERENCES

- [1]. A.V Umarov, B.A Mirsalixov, D.K Djumabayev, F.X Xusnuddinov *Development and research of layered composite materials based on Cu₂ZnSnS (Se) 4 for solar cells. Volume 401, V International Scientific Conference "Construction Mechanics, Hydraulics and Water Resources Engineering" (CONMECHYDRO - 2023)*
- [2]. Zikrillae, N. F., & Ayupov, K. S. (2022). *Influence of Impurity Clusters on the Electronic Properties of Silicon-based Materials*. Technical Science and Innovation, 4(12)
- [3]. Oganov, A. R., Lyakhov, A. O., & Valle, M. (2011). *How Evolutionary Crystal Structure Prediction Works—and Why*. Accounts of Chemical Research.
- [4]. S.V. Rogozhkin, A.A. Bogachev, A.A. Nikitin, A.L. Vasiliev, M.Yu. Presnyakov, M.Tomut, Ch.Trautmann, TEM analysis of radiation effects in ODS steels induced by swift heavy ions. Nuclear Instruments and Methods in Physics Research B 486 (2021)
- [5]. Clifford E. Kintner, "Free-Space Measurements of Dielectrics and Three-Dimensional Periodic Metamaterials" (2017). Theses and Dissertations. 2557.
- [6]. P.K. Johnston, E. Doyle and R.A. Orzel, «Phenolics: A Literature Review of Thermal Decomposition Products and Toxicity», Journal of the American College of Toxicology Vol.7, №2.
- [7]. Butler, K. T., Davies, D. W., Cartwright, H., & Walsh, A. (2018). *Machine learning for molecular and materials science*. Nature, 559(7715).
- [8]. Karniadakis, G. E., et al. (2021). *Physics-informed machine learning*. Nature Reviews Physics, 3(6), 422-440.