EEE 4709 Project Report

Project Title: Synergistic Convergence of Computational Paradigms and Machine Learning Heuristics for the Prognostication of Perovskite Crystallography, Formability, and Thermodynamic Viability

Group No: A4

Author(s)/Student(s) Names:

- Zerin Yeasmin(108)

- Adib Sadman Al Haque(103)

- Omar Faruk(118)

Institution & Course Name:

- Islamic University of Technology,

- EEE4709 (Artificial Intelligence)

Instructor Name: Md Arefin Rabbi Emon

Submission Date: _____

Contents

ABSTRACT

- Summarize the project's objectives, methodology, key results, and conclusions in 200-300 words.

- Keep it concise yet informative. See the abstracts from some papers to get some idea about how to write this section.

1 INTRODUCTION

1.1 Background and Motivation

Perovskite materials, particularly ABO-type oxides, have gained significant attention due to their unique structural, electrical, and magnetic properties, making them suitable for various applications such as solar cells, fuel cells, and superconductors. However, the traditional trial-and-error approach to discovering new perovskite materials is time-consuming and resource-intensive. Machine learning (ML) offers a promising alternative by enabling the rapid prediction of crystal structures and formability, thereby accelerating the discovery of new materials with desired properties. This project aims to leverage ML techniques to classify crystal structures and predict the formability of ABO perovskites, contributing to the development of more efficient and sustainable materials for renewable energy applications.

1.2 Problem Statement

The discovery and optimization of ABO perovskites for various applications, such as photovoltaics, fuel cells, and electronic devices, remain a significant challenge due to the vast compositional space and structural variability of these materials. Traditional experimental and computational methods for identifying stable and high-performing perovskites are time-consuming, costly, and often limited in predictive accuracy. Despite significant advancements, a major hurdle in perovskite research is accurately predicting the crystal structure, stability, and formability of novel compositions before synthesis. The need for a robust, data-driven approach to classify perovskite structures and assess their viability is crucial for accelerating material discovery and application development. This project aims to address these challenges by leveraging machine learning techniques to predict the formability and crystal structure of ABO perovskites. By integrating insights from multiple research studies, we seek to enhance the predictive capabilities of perovskite material screening, ultimately contributing to the development of high-performance materials for energy and electronic applications.

Existing Gaps in Literature and Practice

1. **Limited Predictive Accuracy of Traditional Methods**

   - Computational approaches such as Density Functional Theory (DFT) are widely used but require high computational resources, making large-scale screening impractical.

   - Empirical rules like the Goldschmidt tolerance factor provide general guidance but often fail to accurately predict perovskite stability for complex compositions.

2. **Inconsistent Data and Lack of Unified Frameworks**

   - Studies on perovskite classification and formability prediction have used different datasets and methodologies, leading to inconsistencies in results.
   - A comprehensive, standardized dataset incorporating various machine learning models is needed to improve prediction reliability.

3. **Challenges in Scaling Up Perovskite Applications**

   - While perovskite solar cells have shown high power conversion efficiencies, issues related to material stability and scalability hinder commercialization.
   - Predicting long-term stability remains difficult due to the complex interplay of structural, electronic, and environmental factors.

4. **Limited Exploration of Double Perovskites**

   - Research on single perovskites (ABO) is extensive, but the potential of double perovskites (AABBO) remains underexplored.
   - Machine learning models trained on single perovskites need to be extended and validated for double perovskites to unlock new possibilities in material design.

By addressing these gaps, this study will develop a robust, machine learning-driven framework for perovskite classification and stability prediction, providing valuable insights for researchers and industry professionals working in advanced materials and energy solutions.

1.3   Objectives

This project aims to leverage machine learning techniques to enhance the prediction, classification, and stability assessment of ABO perovskites. The key objectives are as follows:

1. **Develop a Machine Learning Model for Perovskite Classification**

   (a) Implement a predictive framework to classify ABO perovskites into different crystal structures (cubic, tetragonal, orthorhombic, rhombohedral).

   (b) Compare multiple machine learning algorithms (e.g., Random Forest, Light-GBM, XGBoost) to determine the most effective model for crystal structure classification.

2. **Predict the Formability of Perovskite Structures**

   (a) Utilize machine learning to predict whether a given ABO composition can form a stable perovskite structure.

   (b) Identify key features influencing formability, such as the Goldschmidt tolerance factor, octahedral factor, electronegativity, and ionic radii.

3. **Assess the Stability of ABO and Double Perovskites**

(a) Integrate thermodynamic stability analysis with machine learning predictions to identify the most stable perovskite compositions.

(b) Extend the model to predict the formability and stability of double perovskites (AABBO), expanding the chemical space explored.

4. **Create a Unified and Standardized Dataset**

(a) Combine data from existing research reports and additional sources to build a comprehensive dataset for perovskite classification and prediction.

(b) Perform feature selection and correlation analysis to refine input parameters for machine learning models.

5. **Validate the Model's Accuracy and Performance**

(a) Conduct k-fold cross-validation to ensure the robustness and generalizability of the machine learning predictions.

(b) Compare model predictions with existing experimental and computational results to verify accuracy.

6. **Provide Insights for Perovskite-Based Applications**

(a) Identify promising perovskite compositions for solar cells, fuel cells, and electronic applications.

(b) Offer recommendations for experimental synthesis and further optimization based on model predictions.

Expected Outcomes By the end of this study, the following will be achieved:

- A validated machine learning model for ABO perovskite classification and formability prediction.

- A refined dataset integrating multiple research sources for enhanced prediction accuracy.

- Identification of stable perovskite compositions suitable for renewable energy applications.

- Insights into double perovskite structures and their potential advantages.

- A scalable and efficient approach for accelerating perovskite materials discovery.

This research will contribute to the development of advanced materials for energy and electronic applications, reducing the reliance on costly trial-and-error experimental methods.

1.4 Scope and Limitations

Scope: The project focuses on ABO-type perovskites and uses publicly available datasets for training and validation. The model will be evaluated using standard performance metrics such as accuracy, precision, recall, and F1 score.

Limitations: The model's performance is limited by the quality and quantity of available data. Additionally, the model assumes that the perovskites are in their ground state and does not account for temperature or pressure effects.

2 LITERATURE REVIEW / RELATED WORK

2.1 Existing Studies

Several studies have explored the use of machine learning for predicting perovskite properties. For example, **Behara et al.** used the Light Gradient Boosting Machine (Light GBM) algorithm to classify ABO perovskites into different crystal structures with an accuracy of 80.3%. **Ahmad et al.** achieved 90.53% accuracy in crystal structure classification and 98.57% accuracy in formability prediction using the Random Forest algorithm. **Talapatra et al.** focused on the formability and thermodynamic stability of single and double perovskites, identifying 414 promising candidates for experimental synthesis.

2.2 Comparison with Existing Work

This project builds upon the work of **Behara et al.** and **Ahmad et al.** by combining their approaches to improve the accuracy of crystal structure classification and formability prediction. Unlike previous studies, this project also incorporates SHapley Additive exPlanations (SHAP) analysis to identify the most influential features, providing deeper insights into the factors that determine perovskite stability and formability.

3 SYSTEM ARCHITECTURE / EXPERIMENTAL SETUP

3.1 OVERALL SYSTEM DESIGN/MODEL DESCRIPTION

Proposed Model

- The machine learning model will classify ABO perovskites into different crystal structures (cubic, tetragonal, etc.) and predict their formability.

- Algorithms like Random Forest, XGBoost, and LightGBM will be compared to determine the most accurate model.

- SHAP analysis will be used to interpret feature importance.

3.2 HARDWARE AND SOFTWARE REQUIREMENTS

Hardware

- GPU-enabled computing system (for faster model training, if needed)

- Standard CPU-based systems for model development

- High-storage capacity for dataset processing

Software

- **Programming Language**: Python

- **Libraries**: NumPy, Pandas, Scikit-learn, TensorFlow, SHAP

- **Visualization Tools**: Matplotlib, Seaborn

- **Database & Data Handling**: SQLite, CSV/Excel files

- **Web Application:** Marimo

3.3   DATA SOURCES AND PREPROCESSING

Dataset Sources

- Publicly available datasets on ABO perovskites from research papers and repositories

- Experimental data from previous literature (Behara et al., Ahmad et al., Talapatra et al.)

Data Preprocessing Steps

1. **Handling Missing Values**: Impute missing values using mean/mode.

2. **Normalization & Scaling**: Standardize numerical features like electronegativity and ionic radii.

3. **Feature Encoding**: Convert categorical variables (e.g., crystal structures) into numerical form.

Feature Selection: Use correlation analysis and RFECV to identify the most relevant features.

4   METHODOLOGY

4.1   THEORETICAL FOUNDATIONS

Key Concepts

- **Goldschmidt Tolerance Factor ()**

- **Octahedral Factor ()**

- **Machine Learning Algorithms**

    - Random Forest: Ensemble learning method using decision trees
    - XGBoost: Gradient boosting decision trees optimized for performance
    - SVM: Support Vector Machine Classification
    - LightGBM: Efficient boosting method for large datasets

4.2 EXPERIMENTAL SETUP / ALGORITHM

Steps:

1. **Data Collection**

2. **Preprocessing & Feature Engineering**

3. **Model Selection & Training**

4. **Hyperparameter Tuning**

5. **Validation & Performance Evaluation**

6. **RFECV Analysis for Feature Importance**

7. **Comparison with Experimental Data**

8. **Application in Perovskite Research**

   Pseudocode:

- *Load dataset*

- *Preprocess data (normalize, encode, handle missing values)*

- *Split data into training and test sets*

- *For each ML model (Random Forest, SVM, XGBoost, LightGBM):*

- *Train model on training set*

- *Validate using cross-validation*

- *Evaluate performance using accuracy, F1-score*

- *Perform RFECV analysis*

- *Select best-performing model*

- *Predict formability and stability of new perovskites*

4.3 ASSUMPTIONS AND CONSTRAINTS

Assumptions

- The dataset used is representative of real-world perovskite compositions.

- The model predictions are based on the assumption that perovskites are in their ground state.

- Environmental factors like temperature and pressure are not explicitly considered.

Constraints

- Limited availability of high-quality experimental datasets.

- Performance depends on feature engineering and dataset quality.

- Predictions require experimental validation for real-world applications.

5 RESULTS AND ANALYSIS

5.1 PERFORMANCE METRICS

- Summarize previous research or methodologies related to the topic.

- Discuss strengths and limitations of existing approaches.

5.2 EXPERIMENTAL RESULTS

- Highlight how this project differs from or builds upon prior studies.

- Justify why a new approach is needed (if applicable).

5.3 CHALLENGES AND ERROR ANALYSIS

- Highlight how this project differs from or builds upon prior studies.

- Justify why a new approach is needed (if applicable).

6 DISCUSSION AND INSIGHTS

6.1 CRITICAL EVALUATION

- Summarize previous research or methodologies related to the topic.

- Discuss strengths and limitations of existing approaches.

6.2 PRACTICAL APPLICATIONS

- Highlight how this project differs from or builds upon prior studies.

- Justify why a new approach is needed (if applicable).

7 FUTURE WORK AND IMPROVEMENTS

7.1 POSSIBLE ENHANCEMENTS

- Summarize previous research or methodologies related to the topic.

- Discuss strengths and limitations of existing approaches.

7.2 SCALABILITY AND DEPLOYMENT

- Highlight how this project differs from or builds upon prior studies.

- Justify why a new approach is needed (if applicable).

7.3 POTENTIAL RESEARCH DIRECTIONS

- Highlight how this project differs from or builds upon prior studies.

- Justify why a new approach is needed (if applicable).

8 ETHICAL CONSIDERATIONS AND SUSTAINABILITY

8.1 ETHICAL ISSUES

While this project focuses on machine learning (ML)–based prediction of perovskite materials, several ethical concerns must be considered, particularly regarding data integrity, bias, environmental impact, and transparency.

8.1.1 1. Data Privacy and Security

- Concern: The study relies on large datasets compiled from experimental and computational sources, some of which may be proprietary or restricted.

- Ethical Implication: Ensuring that data is obtained from open-access or properly licensed sources is crucial to prevent intellectual property violations.

- Mitigation: Only publicly available datasets and properly cited research will be used.

8.1.2 2. Algorithmic Bias and Model Fairness

- Concern: Machine learning models can inherit biases from training data, leading to skewed or unreliable predictions.

- Example: If the dataset primarily consists of perovskites with specific cation combinations, the model may struggle to generalize for underrepresented compositions.

- Ethical Implication: Biased models can mislead researchers, causing them to overlook potentially promising perovskite compositions.

- Mitigation:

- Using diverse and well-balanced datasets from multiple sources.

- Applying bias detection techniques to ensure the model does not disproportionately favor certain compositions.

8.1.3 3. Environmental and Sustainability Concerns

- Concern: Some perovskite materials contain toxic elements (e.g., lead in Pb-based perovskites), raising concerns about environmental contamination.

- Ethical Implication: Promoting perovskites without considering their environmental impact could lead to harmful waste and pollution.

8.1.4 4. Transparency and Reproducibility

- Concern: Black-box machine learning models may lack interpretability, making it difficult for scientists to understand or verify predictions.

- Ethical Implication: If researchers blindly trust ML-generated predictions without validation, it could lead to misleading conclusions.

8.1.5 5. Responsible AI and Scientific Integrity

- Concern: The results of this study may influence investment, funding, or research directions in material science. If misused, misleading or exaggerated claims could misguide stakeholders.

- Ethical Implication: Misrepresentation of ML capabilities in material discovery could erode trust in AI-driven research.

- Mitigation:

- Clearly stating limitations and uncertainties in the model's predictions.

- Encouraging experimental validation of AI-generated perovskite predictions before real-world implementation.

8.2 SUSTAINABILITY

This project has significant long-term implications for environmental sustainability, economic viability, and societal benefits by accelerating the discovery of high-performance perovskite materials. The integration of machine learning (ML) can lead to more efficient, eco-friendly, and cost-effective materials for energy, electronics, and industrial applications.

8.2.1 1. Environmental Impact

8.2.2 Renewable Energy and Carbon Reduction

- Perovskite solar cells (PSCs) have demonstrated power conversion efficiencies (PCEs) exceeding 25%, making them a strong alternative to traditional silicon-based solar cells.

- Increased adoption of PSCs can significantly reduce dependence on fossil fuels, lowering CO emissions and mitigating climate change.

Toxicity and Waste Management

- Some perovskite materials contain lead (Pb), which poses environmental hazards.

- This project prioritizes the prediction and discovery of lead-free perovskites, reducing toxic waste and promoting eco-friendly alternatives.

Sustainable Manufacturing

- Traditional material discovery is resource-intensive and involves excessive chemical synthesis and experimental trials.

- ML-driven perovskite discovery minimizes experimental waste by identifying promising materials before physical testing, leading to greener lab practices.

8.2.3   2. Economic Viability

Lower Production Costs for Solar Cells

- Perovskites are cheaper to manufacture than silicon, requiring lower energy input and simpler fabrication processes.

- If successfully commercialized, perovskite-based solar panels could lower electricity costs, making renewable energy more accessible worldwide.

Job Creation in Green Technology

- Advances in perovskite research can drive growth in clean energy industries, creating jobs in solar panel manufacturing, energy storage, and material science.

- Expansion into electronic and optoelectronic applications could further boost economic development.

Faster Material Commercialization

- Traditional perovskite research takes years due to trial-and-error synthesis.

- Machine learning accelerates material discovery, shortening the time needed to bring new perovskites to market, benefiting industries from renewable energy to semiconductors.

8.2.4   3. Societal Impact

Energy Access for Remote Areas

- Lightweight, flexible perovskite solar cells can provide off-grid energy solutions, improving electricity access in rural and underdeveloped regions.

- Affordable solar solutions contribute to energy equity, improving quality of life in low-income communities.

Advancing Scientific Innovation

- By combining machine learning with materials science, this project supports interdisciplinary research, fostering new breakthroughs in AI-driven material discovery.

- Open-access datasets and ML models can be shared globally, allowing researchers worldwide to contribute to and benefit from perovskite advancements.

Reduction in Electronic Waste (E-Waste)

- Perovskite materials can be engineered for sustainable electronics, replacing traditional materials in batteries, displays, and semiconductors.

- If designed for recyclability, perovskites could reduce electronic waste, promoting a circular economy in the tech industry.

9   CONCLUSION

References

[1] S. Behara, T. Poonawala, and T. Thomas, "Crystal structure classification in ABO_3 perovskites via machine learning," *Computational Materials Science*, vol. XXX, pp. 1–10, 2020, doi:10.1016/j.commatsci.2020.110191.

[2] M. U. Ahmad *et al.*, "ABO_3 perovskites' formability prediction and crystal structure classification using machine learning," *arXiv preprint*, 2022, arXiv:2202.10125.

[3] A. Talapatra *et al.*, "A machine learning approach for the prediction of formability and thermodynamic stability of single and double perovskite oxides," *Chemistry of Materials*, vol. 33, no. 3, pp. 845–858, 2021, doi:10.1021/acs.chemmater.0c03402.

[4] NREL, "Best Research-Cell Efficiency Chart," *National Renewable Energy Laboratory (NREL)*, 2023. [Online]. Available: https://www.nrel.gov/pv/cell-efficiency.html.

[5] T. M. Bartel *et al.*, "New tolerance factor to predict the stability of perovskite oxides and halides," *Science Advances*, vol. 5, no. 2, p. eaav0693, 2019, doi:10.1126/sciadv.aav0693.

[6] Materials Project Database, "Perovskite Oxide Dataset," *Lawrence Berkeley National Laboratory*, 2022. [Online]. Available: https://materialsproject.org/.

[7] J. Xu *et al.*, "The role of the Goldschmidt tolerance factor in the structure and stability of perovskite oxides: A machine learning approach," *Journal of Materials Chemistry A*, vol. 9, no. 10, pp. 5532–5542, 2021, doi:10.1039/D0TA09647H.

[8] G. Kieslich, S. Sun, and A. K. Cheetham, "Solid-state principles applied to organic–inorganic perovskites: new tricks for an old dog," *Chemical Science*, vol. 5, no. 12, pp. 4712–4715, 2014, doi:10.1039/C4SC02211D.

[9] C. Sun *et al.*, "Toward a universal model for predicting the formability of perovskite materials," *Physical Review Materials*, vol. 5, no. 8, p. 085801, 2021, doi:10.1103/PhysRevMaterials.5.085801.

[10] S. Das, B. Maiti, and T. Roy, *Perovskite Materials for Energy and Electronics: Advances and Applications*, 1st ed. Springer, 2022, doi:10.1007/978-3-030-87660-3.