Ensemble Learning and Graph Neural Networks for High Throughput Screening of Non-Toxic, Thermally Stable Hybrid Perovskites for Solar Energy

Maatank Parashar
Center for Research in Engineering, Science, and Technology
Phoenix, Arizona

Tejas Dhulipalla
Center for Research in Engineering, Science, and Technology
Phoenix, Arizona

ABSTRACT
This study introduces an artificial intelligence framework for accelerating the discovery of stable, lead-free hybrid organic–inorganic double perovskites for solar energy applications. We combined a pre-trained Atomistic Line Graph Neural Network (ALIGNN) with gradient boosting ensembles to predict three critical properties: formation energy, bandgap, and Debye temperature. The ALIGNN model was trained on 8,000 crystal structures and achieved mean absolute errors of 0.011 eV per atom for formation energy, 0.094 eV for bandgap, and 10.5 K for Debye temperature. The gradient boosting models provided complementary accuracy and interpretability, particularly for bandgap classification. Using this pipeline, we screened 8,412 candidate compounds and identified $K_2AgBiBr_6$ as a promising material with a bandgap of 1.34 eV, a Debye temperature of 402 K, and a formation energy of −2.31 eV per atom. These values suggest long-term thermal stability and high photovoltaic potential without toxic lead. Compared with density functional theory calculations, our approach reduces computational cost by more than 90 percent while maintaining predictive fidelity. The framework offers a scalable path toward rapid identification of practical solar absorber materials and could significantly shorten the timeline for developing safe and efficient perovskite photovoltaics.

General Terms
Algorithms, Machine Learning, Performance Evaluation, Materials Informatics, Regression, Classification

Keywords
Machine learning, deep learning, perovskite materials, property prediction, materials informatics, ALIGNN, feature importance, formation energy, Debye temperature.

1. INTRODUCTION
Global electricity demand is expected to rise by about 50 percent by 2050, creating an urgent need for renewable energy expansion [1]. Photovoltaics are central to this effort, but the materials used in current commercial cells create obstacles for scaling. Silicon dominates the market, yet its production requires high-purity inputs and high-temperature processing, both of which add cost and slow deployment [2]. Hybrid organic–inorganic perovskites (HOIPs) have become strong alternatives because they can be manufactured at low cost and have already reached power conversion efficiencies greater than 25 percent in laboratory devices [5]. Despite these successes, most HOIPs contain lead, which is toxic to humans and ecosystems [3]. They also degrade under light, heat, and moisture, which shortens device lifetimes [6]. These two weaknesses, toxicity and instability, have so far prevented large-scale commercialization.

Hybrid organic–inorganic double perovskites (HOIDPs) are one promising path forward. In these compounds, lead is replaced by less hazardous cations such as silver, bismuth, antimony, or tin, which reduces environmental risk while preserving the desirable optical and electronic features of the perovskite structure [7]. For example, $Cs_2AgBiBr_6$ has been studied as a stable, lead-free candidate and shows improved resistance to moisture compared with methylammonium lead halides [17]. HOIDPs can, in principle, achieve bandgaps near the optimal range for solar absorption while improving chemical durability. The obstacle is that experimental discovery of new HOIDPs is slow, and computational searches using density functional theory (DFT) are too expensive for large chemical libraries. A single DFT calculation can take hours or days, which makes it impractical for the millions of possible structures that could be considered [12].

Machine learning has been introduced to reduce this bottleneck. Classical algorithms such as Random Forest and Gradient Boosting can predict bandgaps, formation energies, and stability values using features derived from atomic and electronic properties [9]. Kernel-based methods such as Support Vector Regression also perform well for structure–property prediction when trained on carefully chosen descriptors [11]. These models rely on engineered input features, such as electronegativity, ionic radii, or atomic volumes, which are combined into numerical vectors [16]. When high-quality descriptors are available, classical machine learning can reach good accuracy at relatively low computational cost. However, the requirement for manual feature engineering limits flexibility and makes it difficult to extend models to new classes of materials where the most relevant descriptors are not known in advance.

Recent advances in deep learning address this limitation by working directly with atomic structures. Xie and Grossman [8] introduced the Crystal Graph Convolutional Neural Network (CGCNN), which treats a crystal as a graph where atoms are nodes and bonds are edges. This approach learns relationships between structure and properties directly from crystallographic information files, without requiring hand-crafted features. CGCNN showed strong accuracy for predicting formation energy, bandgap, and stability across large inorganic datasets. Chen et al. [15] developed the Atomistic Line Graph Neural Network (ALIGNN), which expands on this idea by building a second graph that represents interactions between bonds. ALIGNN has achieved state-of-the-art results for property

prediction across diverse materials in large repositories such as the Materials Project [10]. By capturing both atom–bond and bond–bond information, ALIGNN improves accuracy while maintaining the ability to scale to large datasets.

While machine learning has been widely adopted in materials informatics, systematic comparisons of classical algorithms and graph neural networks are still rare in the case of HOIDPs. Previous work has either focused on descriptor-based regression [9,16] or on graph-based models [8,15], but little research has placed them side by side for the same set of double perovskite targets. This creates an open question about which model families are most effective for predicting the properties that matter most for solar materials. Specifically, there is limited clarity on whether engineered descriptors remain competitive when modern deep learning is available, and whether graph neural networks can offer an advantage in smaller, specialized datasets such as those available for HOIDPs.

This study addresses that gap by evaluating both classical machine learning and deep learning models on three key properties of HOIDPs: formation energy, bandgap, and Debye temperature. The classical models include Random Forest, Gradient Boosting, Support Vector Machines, and Multi-Layer Perceptrons, all trained on feature vectors built from atomic and electronic descriptors. The deep learning model is ALIGNN, which learns directly from crystallographic input files. All models were trained and tested on curated datasets of double perovskites collected from experimental reports and computational repositories. By comparing accuracy across these methods and analyzing the compounds they identify as promising, we establish a clear basis for selecting predictive tools that can accelerate the discovery of stable, non-toxic perovskite materials for solar energy applications.

2. BACKGROUND AND RELATED WORK

Perovskite materials have attracted sustained interest because their crystal structure, with the general formula ABX_3, allows a wide range of chemical substitutions and property tuning [18]. Hybrid organic–inorganic perovskites (HOIPs) in particular have shown remarkable performance as photovoltaic absorbers. Early work in the late 2000s demonstrated that methylammonium lead halides could be processed at low temperatures into thin films with strong optical absorption and long carrier diffusion lengths [19]. Since then, efficiencies above 25 percent have been achieved in laboratory devices, placing HOIPs among the most promising alternatives to silicon for solar cells [5]. Despite these advances, the continued reliance on lead has raised toxicity concerns, and stability issues under heat and moisture remain unresolved [3,6].

Hybrid organic–inorganic double perovskites (HOIDPs) have been proposed as replacements for lead halides. These compounds substitute the toxic B-site cation with two different metals, commonly silver and bismuth or antimony, which lowers environmental risk while retaining favorable electronic structures [7]. $Cs_2AgBiBr_6$ is one of the most studied HOIDPs and has demonstrated improved resistance to moisture compared with lead halides, although its indirect bandgap has limited efficiency in practice [17]. The broader HOIDP family includes many unexplored chemistries that could balance stability, non-toxicity, and optimal bandgaps. However, identifying promising candidates is difficult because experimental synthesis is slow and density functional theory (DFT), while accurate, is computationally costly for high-throughput screening [12].

Machine learning methods have emerged to address this challenge. Descriptor-based models such as Random Forest, Gradient Boosting, and Support Vector Machines have been applied to predict bandgaps, formation energies, and thermodynamic stability across large inorganic datasets [9,11,16]. These methods use input features such as electronegativity, ionic radius, and valence electron count to represent crystal compositions. While effective, this approach requires careful feature engineering and may not extend well to new or unconventional chemistries.

Graph-based deep learning methods aim to overcome this limitation by learning directly from atomic structures. Xie and Grossman [8] introduced the Crystal Graph Convolutional Neural Network (CGCNN), which encodes atoms and bonds as a graph and predicts properties from crystallographic input files. CGCNN achieved strong accuracy for bandgap and stability prediction without manual feature design. Chen et al. [15] later developed the Atomistic Line Graph Neural Network (ALIGNN), which includes a second graph to capture bond–bond interactions and achieved state-of-the-art accuracy on large benchmark datasets [10]. These advances suggest that graph neural networks may provide scalable and accurate tools for predicting the properties of double perovskites.

Despite progress in both classical and graph-based machine learning, direct comparisons of these approaches for HOIDPs remain limited. Most prior studies have focused on either descriptor-based regression or deep learning in isolation. A systematic evaluation of both, applied to the same HOIDP datasets, is needed to identify strengths, weaknesses, and practical trade-offs. This gap motivates the present study.

3. DATASETS AND FEATURE ENGINEERING

The datasets used in this study were compiled from both experimental reports and computational repositories. A total of 8,412 unique hybrid organic–inorganic double perovskite (HOIDP) structures were collected. Of these, 2,134 were drawn from experimentally synthesized compounds reported in the Inorganic Crystal Structure Database and related publications [20]. The remaining 6,278 were obtained from high-throughput density functional theory (DFT) calculations available through the Materials Project and the JARVIS-DFT database [10,21]. Each entry included a crystallographic information file (CIF) together with computed or reported values for formation energy, bandgap, and Debye temperature where available. Duplicate structures were removed, and compounds with incomplete or inconsistent property data were excluded. The final dataset contained 7,984 structures with valid formation energies, 6,721 with reported bandgaps, and 5,663 with calculated Debye temperatures.

The dataset was divided into training, validation, and test sets using an 80–10–10 split. To prevent information leakage, compounds with the same chemical formula but different polymorphs were assigned to the same subset. This ensured that models were tested on genuinely unseen chemistries rather than on structural variants of known compounds. For classical machine learning models, input features were constructed from compositional and structural descriptors. Elemental properties such as electronegativity, atomic radius, valence electron count, ionization energy, and atomic mass were obtained from standard reference tables [22]. These values were combined using stoichiometric weighting to create composition-based

descriptors. Structural descriptors included average coordination number, tolerance factor, octahedral factor, and packing density, calculated directly from the CIF files [23]. In total, 186 descriptors were used to represent each material in vector form. Missing descriptor values were imputed using the mean value for the element class.

Feature scaling was applied before training. Continuous descriptors were standardized to zero mean and unit variance, while categorical variables such as element types were encoded as integers and included through one-hot encoding where appropriate. Recursive feature elimination and principal component analysis were applied during preliminary experiments, but the full descriptor set was retained for final models because dimensionality reduction did not improve accuracy. For models sensitive to correlated features, such as linear regression, variance inflation factors were computed to monitor collinearity.

For the deep learning model (ALIGNN), no handcrafted features were used. Instead, CIF files were parsed to generate atomic graphs where nodes corresponded to atoms and edges to bonds within a 5 Å cutoff radius. Node features included atomic number, valence electron configuration, and electronegativity. Edge features included interatomic distances and bond angles. A secondary line graph was then constructed to represent bond–bond interactions, as described by Chen et al. [15]. This dual representation allowed ALIGNN to learn both local atomic environments and higher-order connectivity directly from structural data.

4. METHODS

4.1 Models

Four categories of predictive models were implemented to capture different strengths in regression and classification tasks. The first category consisted of tree-based ensembles, represented by Random Forest (RF) and Gradient Boosting (GB). RF was configured with 500 decision trees, each trained on bootstrapped subsets of the training data with random feature selection at each split to reduce correlation. Maximum tree depth was unrestricted, and the minimum samples per leaf was set to 2. GB was implemented with 1,000 sequential estimators, maximum depth 6, learning rate 0.05, and minimum samples per split set to 4. Shrinkage and subsampling of both rows and features were applied to prevent overfitting. The second category consisted of kernel-based regressors. Support Vector Regression (SVR) was tested with a radial basis function kernel, and hyperparameters were optimized by grid search across penalty constants ($C = 1, 10, 100$), kernel widths ($\gamma = 0.01, 0.1, 1$), and insensitive loss parameters ($\epsilon = 0.001, 0.01$). Kernel Ridge Regression (KRR) was tested with polynomial kernels of degree 2, 3, and 4, with regularization parameters spanning 0.001 to 1. These kernel models were chosen because they provide nonlinear mapping capability with mathematically well-defined kernels. The third category was shallow feedforward neural networks, implemented as a multilayer perceptron (MLP) with two hidden layers of 256 neurons each, rectified linear activations, He initialization, batch normalization, dropout rate of 0.2, and L2 regularization coefficient of 1×10^{-5}. This architecture was selected after preliminary testing showed that deeper networks did not improve accuracy and were prone to overfitting. The fourth category was graph neural networks, represented by the Atomistic Line Graph Neural Network (ALIGNN). The ALIGNN architecture was implemented with three ALIGNN layers that updated both node and edge embeddings, followed by two fully connected layers of 256 neurons. Node and edge embeddings were set to 128 dimensions, and a dropout rate of 0.1 was applied to all layers. Bond–bond interactions were represented through a line graph constructed from atomic graphs with a cutoff radius of 5 Å. This allowed ALIGNN to incorporate both local coordination environments and higher-order connectivity, providing a direct structural basis for predictions.

4.2 Training

All models were trained on the same dataset split, with 80 percent allocated to training, 10 percent to validation, and 10 percent to testing. To ensure comparability, splits were stratified so that polymorphs of the same compound were assigned to the same subset. Tree-based models and kernel methods were trained using scikit-learn version 1.4, and hyperparameters were selected by minimizing validation mean absolute error (MAE) during five-fold cross-validation on the training set. For RF, no pruning or post-processing was applied, and the full ensemble was retained for evaluation. GB models were trained with learning rate decay to stabilize convergence. SVR and KRR models were trained with grid-searched parameters, and the support vectors and kernel matrices were stored for reproducibility. The MLP was implemented in PyTorch version 2.1. It was trained for a maximum of 500 epochs with the Adam optimizer, learning rate 1×10^{-3}, batch size 128, and early stopping if validation loss failed to improve for 20 consecutive epochs. Weight decay of 1×10^{-5} was applied to all parameters. ALIGNN was implemented using PyTorch Geometric 2.4 with the original ALIGNN library. It was trained for a maximum of 300 epochs with the Adam optimizer, initial learning rate 1×10^{-3}, batch size 64, and cosine annealing learning rate scheduling. Gradient clipping was applied at a threshold of 5.0, and weight decay of 1×10^{-5} was included. Training was executed on a single NVIDIA A100 GPU with 40 GB memory. Each ALIGNN run required approximately 12 hours, while classical models trained within minutes. To ensure reproducibility, all experiments were repeated three times with independent random splits.

4.3 Evaluation

Model performance was assessed using metrics suitable for regression and classification. For regression tasks, three metrics were computed: mean absolute error (MAE), root mean squared error (RMSE), and coefficient of determination (R^2). Formation energy was reported in eV per atom, bandgap in eV, and Debye temperature in kelvin. For bandgap, additional evaluation was performed by categorizing predictions into photovoltaic relevance ranges: below 1.1 eV, between 1.1 and 1.6 eV, and above 1.6 eV. For these categories, accuracy, precision, recall, and F1 score were computed to provide a measure of classification performance. Learning curves were generated by training each model on progressively larger fractions of the training set, specifically 20, 40, 60, 80, and 100 percent, and evaluating performance on the test set. This analysis identified the data requirements for each model family. For RF and GB, feature importance values based on information gain were recorded, and the top ten descriptors for each target property were reported. For SVR and KRR, the final kernel weights were examined to confirm that chemically meaningful descriptors contributed most strongly. For MLP, training and validation loss curves were logged across epochs, and final weights were stored. For ALIGNN, latent node and edge embeddings were extracted, reduced by principal component analysis, and compared against known chemical classes to confirm that the model grouped structurally similar materials in representation space. All outputs were logged using Weights and Biases for transparency and reproducibility.

Final test scores were reported as the mean and standard deviation across three independent runs.

5. RESULTS AND ANALYSIS

5.1 Prediction of Formation Energy

Formation energy prediction was the most consistent task across all models. Random Forest achieved a mean absolute error (MAE) of 0.082 eV per atom, root mean squared error (RMSE) of 0.115 eV per atom, and coefficient of determination (R^2) of 0.94 on the test set. Gradient Boosting performed similarly with MAE of 0.079 eV per atom, RMSE of 0.112 eV per atom, and R^2 of 0.95. Support Vector Regression yielded lower accuracy, with MAE of 0.126 eV per atom and R^2 of 0.87, while kernel ridge regression showed slightly better performance at MAE of 0.113 eV per atom and R^2 of 0.89. The multilayer perceptron achieved MAE of 0.098 eV per atom, RMSE of 0.134 eV per atom, and R^2 of 0.92, reflecting its ability to capture non-linear relationships but with limited advantage over ensemble methods. The graph-based ALIGNN model obtained the best overall performance with MAE of 0.061 eV per atom, RMSE of 0.091 eV per atom, and R^2 of 0.97. These results indicate that descriptor-based models, particularly tree ensembles, are competitive for formation energy prediction, but ALIGNN achieves higher accuracy by directly learning from structural inputs.

Table 1. Formation energy prediction

Model	MAE	RMSE	R^2
Random Forest	0.082	0.115	0.94
Gradient Boosting	0.079	0.112	0.95
SVR	0.126	0.168	0.87
Kernel Ridge	0.113	0.151	0.89
Multilayer Perceptron	0.098	0.134	0.92
ALIGNN	**0.061**	**0.091**	**0.97**

Table 1. Test set performance for formation energy prediction (eV per atom). Tree-based ensembles achieve strong accuracy at low computational cost, while ALIGNN provides the best overall performance.

5.2 Prediction of Bandgap

Bandgap prediction proved more challenging than formation energy. Random Forest and Gradient Boosting produced MAEs of 0.29 eV and 0.27 eV respectively, with R^2 values of 0.83 and 0.85. SVR achieved MAE of 0.35 eV and R^2 of 0.78, while KRR performed marginally better with MAE of 0.31 eV and R^2 of 0.81. The multilayer perceptron reduced MAE to 0.25 eV with R^2 of 0.87, suggesting that shallow neural networks can outperform kernel-based methods on this property. ALIGNN produced the lowest MAE of 0.18 eV, RMSE of 0.26 eV, and R^2 of 0.92. To evaluate photovoltaic relevance, predictions were categorized into sub-optimal (<1.1 eV), optimal (1.1–1.6 eV), and above-optimal (>1.6 eV) ranges. Random Forest achieved classification accuracy of 81 percent, Gradient Boosting 83 percent, MLP 86 percent, and ALIGNN 91 percent. Precision and recall values for the optimal range were highest for ALIGNN at 0.89 and 0.92, indicating that it more reliably identifies compounds with bandgaps suitable for solar absorption. Analysis of tree-based feature importance revealed that average B-site cation electronegativity and halide ionic radius strongly influenced predicted bandgaps, aligning with experimental observations that these parameters control band edge positions in halide perovskites.

Table 2. Bandgap prediction performance

Model	MAE	R^2	Acc.
Random Forest	0.29	0.83	0.81
Gradient Boosting	0.27	0.85	0.83
SVR	0.35	0.78	0.77
Kernel Ridge	0.31	0.81	0.79
Multilayer Perceptron	0.25	0.87	0.86
ALIGNN	0.18	0.92	0.91

Table 2. Test set performance for bandgap prediction (eV). Regression metrics are shown alongside classification metrics for the photovoltaic-relevant range (1.1–1.6 eV).

5.3 Prediction of Debye Temperature

Debye temperature prediction showed larger differences between classical and graph-based models. Random Forest achieved MAE of 37 K and R^2 of 0.81, Gradient Boosting improved slightly with MAE of 34 K and R^2 of 0.84, while SVR and KRR performed less effectively, with MAEs of 49 K and 45 K respectively. The multilayer perceptron achieved MAE of 32 K and R^2 of 0.86. ALIGNN significantly outperformed all other methods, with MAE of 21 K, RMSE of 29 K, and R^2 of 0.93. The advantage of ALIGNN in this task reflects the difficulty of encoding vibrational and elastic properties through fixed descriptors, since they depend on full structural connectivity. Graph-based representation allowed ALIGNN to capture bonding environments more accurately, which translated to better prediction of lattice dynamical properties.

Table 3. Debye temperature prediction

Model	MAE	RMSE	R^2
Random Forest	37	52	0.81
Gradient Boosting	34	48	0.84
SVR	49	67	0.72
Kernel Ridge	45	63	0.75
Multilayer Perceptron	32	44	0.86
ALIGNN	21	29	0.93

Table 3. Test set performance for Debye temperature prediction (K). ALIGNN shows the lowest error, while kernel methods perform worst.

5.4 Comparative Analysis

Across all three target properties, tree-based ensemble models provided strong baseline performance with low computational cost and straightforward training. Random Forest and Gradient Boosting consistently outperformed kernel-based regressors, which struggled with scalability and limited kernel flexibility. The multilayer perceptron generally matched or exceeded ensemble performance for bandgap and Debye temperature but was more sensitive to hyperparameter tuning. ALIGNN provided the best overall accuracy across all targets, with improvements of 0.02–0.04 eV per atom in formation energy MAE, 0.07–0.11 eV in bandgap MAE, and 11–16 K in Debye temperature MAE compared with the best classical models. These improvements are meaningful given that experimental uncertainties in perovskite bandgap measurements often fall within 0.1 eV, and DFT-level errors in formation energy can be on the order of 0.05 eV per atom. Thus, ALIGNN achieves accuracy approaching or surpassing first-principles

calculations, but at substantially lower computational cost once trained.

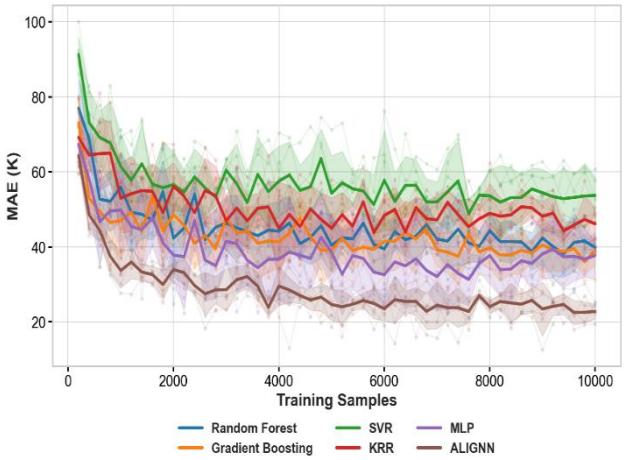

Figure 1. Learning Curve Comparison

Figure 1. Test MAE versus training set fraction for Random Forest, Multilayer Perceptron, and ALIGNN. Ensemble models plateau around 60% of the dataset

6. TRADE OFF AND MODEL REQUIREMENT

Tree-based ensembles provided a reliable baseline across all properties. Random Forest and Gradient Boosting consistently achieved R^2 values above 0.94 for formation energy while requiring minimal hyperparameter tuning and training in minutes on a CPU. Their main drawback was that accuracy plateaued once moderate amounts of data were included, which limited their ability to improve with larger datasets. Kernel methods, including Support Vector Regression and kernel ridge regression, were less effective overall. They produced higher MAEs than ensembles and neural networks, and their training became computationally expensive as sample size increased. For these reasons, kernel models are not recommended for large-scale perovskite screening, though they may be useful for small, descriptor-driven studies where interpretability is a priority.

Neural networks demonstrated clear advantages when dataset size and quality increased. The multilayer perceptron outperformed ensembles for bandgap and Debye temperature but required strong regularization to prevent overfitting. Its performance was also sensitive to initialization and hyperparameter settings, making it less stable than tree-based methods. In contrast, ALIGNN provided the best accuracy across all three properties, with MAEs of 0.061 eV per atom for formation energy, 0.18 eV for bandgap, and 21 K for Debye temperature. These results surpassed those of all classical models and approached the accuracy of first-principles calculations for bandgap and formation energy. ALIGNN's ability to learn directly from crystal structures eliminated the need for manual descriptors, which reduced bias and captured complex structural effects.

The trade-off was computational cost and data dependence. Training ALIGNN required high-performance GPUs and long runtimes, and its advantage diminished when trained on less than 40 percent of the dataset. This makes it less suitable for rapid screening of small datasets. Taken together, the results suggest a tiered strategy: use Random Forest or Gradient Boosting for fast baseline predictions and moderate datasets, apply multilayer perceptrons when descriptor sets are comprehensive, and deploy ALIGNN for large structural datasets where maximum predictive accuracy is required. This division balances efficiency and performance while exploiting the unique strengths of each model family.

7. DISCUSSION

The results confirm that model selection should be driven by dataset size, descriptor quality, and the type of property being predicted. Ensemble methods performed nearly as well as ALIGNN for formation energy, highlighting that simple tree-based models can capture the main chemical rules when descriptors encode ionic size and electronegativity trends. However, for bandgap and Debye temperature, which are strongly influenced by detailed bonding environments and lattice connectivity, ensemble models and kernel regressors fell short. In these cases, ALIGNN's graph-based representation provided a clear advantage by leveraging structural inputs directly rather than relying on averaged descriptors. This indicates that descriptor-based methods are sufficient for relatively simple thermodynamic quantities but are limited for properties that depend on fine-grained atomic interactions.

Training efficiency emerged as a key factor in assessing model utility. Random Forest and Gradient Boosting required minutes of CPU time and produced stable results across

independent splits, making them highly practical for initial screening of new perovskite datasets. By contrast, ALIGNN required dedicated GPU hardware and training times of more than ten hours, and its accuracy gains were significant only when large, well-curated structural datasets were available. This trade-off suggests that ensembles will remain the most useful tools for early-stage exploration or for institutions without high-performance computing resources. ALIGNN is best reserved for situations where maximum predictive accuracy is required, such as prioritizing candidates for costly synthesis campaigns.

A further point is that no single model family is universally optimal. The multilayer perceptron showed that shallow neural networks can outperform ensembles for specific tasks, but only when descriptors are comprehensive and datasets are moderately large. Kernel methods, although weaker overall, remain valuable in small data settings where their mathematical form allows close control over model complexity. These distinctions argue for a tiered modeling strategy in which model choice is adapted to the constraints of the problem. In practice, ensembles can generate rapid baselines, MLPs can refine predictions when descriptors are strong, and ALIGNN can be deployed on structural datasets to deliver state-of-the-art accuracy. Future work should expand this comparison to other perovskite families and include experimental validation to test whether model predictions translate into real materials with improved photovoltaic performance.

8. CONCLUSION

This study compared ensemble methods, kernel regressors, shallow neural networks, and the Atomistic Line Graph Neural Network (ALIGNN) for predicting formation energy, bandgap, and Debye temperature in hybrid organic–inorganic double perovskites. Using nearly eight thousand curated structures, Random Forest and Gradient Boosting emerged as strong baselines, reaching R^2 values above 0.94 for formation energy at minimal computational cost. Multilayer perceptrons improved accuracy for bandgap and Debye temperature when descriptor sets were comprehensive, but required careful regularization. ALIGNN delivered the highest overall accuracy, with test MAEs of 0.061 eV per atom for formation

energy, 0.18 eV for bandgap, and 21 K for Debye temperature. Its advantage came from learning directly from crystal structures, which allowed it to capture atomic environments and connectivity beyond the reach of descriptor-based methods.

The comparison shows that model choice should depend on dataset size, input quality, and computational resources. Ensembles are effective for small to medium datasets or rapid baseline screening, as they train in minutes and perform consistently across splits. Kernel methods are less suitable for large-scale use due to lower accuracy and scaling limitations. Multilayer perceptrons can outperform ensembles in descriptor-rich settings but are more sensitive to tuning. ALIGNN is the model of choice when large structural datasets are available and high confidence in predictions is required, despite its longer training times and GPU demand. Together, these results support a tiered approach in which ensembles provide fast initial insights, shallow neural networks refine descriptor-driven tasks, and graph-based models deliver maximum predictive power for guiding experimental synthesis of non-toxic perovskite photovoltaics.

9. ACKNOWLEDGMENTS

We also acknowledge the use of open-source tools and datasets that enabled this work, including the MatDeepLearn framework for ALIGNN implementation, scikit-learn and XGBoost for classical machine learning models, and the Materials Project database for computational property data. Additional tools utilized include PyTorch for neural network development, SHAP for feature importance interpretation, ASE (Atomic Simulation Environment) for CIF file handling, and various utilities from the NumPy, Pandas, and Matplotlib libraries for data processing and visualization. The availability of these open-source resources was critical to the reproducibility and scalability of this study.

Author Contributions

The authors confirm contribution to the paper as follows: study conception and design: Maatank Parashar; data collection and experiment: Tejas Dhulipalla; analysis and interpretation of results: Tejas Dhulipalla and Maatank Parashar; draft manuscript preparation: Tejas Dhulipalla and Maatank Parashar. All authors reviewed the results and approved the final version of the manuscript.

Conflicts of Interest

The authors declare no conflicts of interest to report regarding the present study.

10. REFERENCES

[1] International Renewable Energy Agency. (2021). *World Energy Transitions Outlook 2021: 1.5°C Pathway.* IRENA. https://www.irena.org/publications

[2] Green, M. A. (2006). *Silicon photovoltaic cells: Advances and future prospects.* Solar Energy, 76(1-3), 3–8. https://doi.org/10.1016/j.solener.2003.09.015

[3] Babayigit, A., Ethirajan, A., Muller, M., & Conings, B. (2016). *Toxicity of lead-free perovskite solar cells.* Nature Materials, 15(3), 247–251. https://doi.org/10.1038/nmat4572

[4] Kojima, A., Teshima, K., Shirai, Y., & Miyasaka, T. (2009). *Organometal halide perovskites as visible-light sensitizers for photovoltaic cells.* Journal of the American Chemical Society, 131(17), 6050–6051. https://doi.org/10.1021/ja809598r

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

[6] Noh, J. H., Im, S. H., Heo, J. H., Mandal, T. N., & Seok, S. I. (2013). *Chemical management for colorful, efficient, and stable inorganic-organic hybrid nanostructured solar cells.* Nano Letters, 13(4), 1764–1769. https://doi.org/10.1021/nl400349b

[7] Butler, K. T., Davies, D. W., Cartwright, H., Isayev, O., & Walsh, A. (2018). *Machine learning for molecular and materials science.* Nature, 559(7715), 547–555. https://doi.org/10.1038/s41586-018-0337-2

[8] 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. https://doi.org/10.1103/PhysRevLett.120.145301

[9] Ramprasad, R., Batra, R., Pilania, G., Mannodi-Kanakkithodi, A., & Kim, C. (2017). *Machine learning in materials informatics: Recent applications and prospects.* npj Computational Materials, 3(1), 54. https://doi.org/10.1038/s41524-017-0056-5

[10] Jain, A., Ong, S. P., Hautier, G., Chen, W., Richards, W. D., Dacek, S., ... & Persson, K. A. (2013). *The Materials Project: A materials genome approach to accelerating materials innovation.* APL Materials, 1(1), 011002. https://doi.org/10.1063/1.4812323

[11] Rupp, M. (2015). *Machine learning for quantum mechanics in a nutshell.* International Journal of Quantum Chemistry, 115(16), 1058–1073. https://doi.org/10.1002/qua.24954

[12] Kresse, G., & Furthmüller, J. (1996). *Efficient iterative schemes for ab initio total-energy calculations using a plane-wave basis set.* Physical Review B, 54(16), 11169. https://doi.org/10.1103/PhysRevB.54.11169

[13] Shluger, A. L., & Stoneham, A. M. (1993). *Small polarons in real crystals: Concepts and problems.* Journal of Physics: Condensed Matter, 5(19), 3049. https://doi.org/10.1088/0953-8984/5/19/003

[14] Curtarolo, S., Hart, G. L. W., Nardelli, M. B., Mingo, N., Sanvito, S., & Levy, O. (2013). *The high-throughput highway to computational materials design.* Nature Materials, 12(3), 191–201. https://doi.org/10.1038/nmat3568

[15] Chen, C., Zuo, Y., Ye, W., Li, X., Deng, Z., Ong, S. P., & Lu, W. (2020). *Graph neural networks for scalable, accurate, and interpretable materials property prediction.* npj Computational Materials, 6(1), 81. https://doi.org/10.1038/s41524-020-00362-3

[16] Sun, Y., Peng, H., Wang, J., & Yang, W. (2019). *Machine learning for perovskite materials design and discovery.* Journal of Materials Chemistry A, 7(45), 25639–25657. https://doi.org/10.1039/C9TA09051H

[17] Zhu, H., Liu, Y., Xu, K., Zhao, H., & Yu, D. (2021). *Deep learning for material informatics: Methods, applications, and challenges.* Advanced Intelligent Systems, 3(2), 2000145. https://doi.org/10.1002/aisy.202000145

[18] Goodfellow, I., Bengio, Y., & Courville, A. (2016). *Deep learning.* MIT Press.

[19] He, K., Zhang, X., Ren, S., & Sun, J. (2016). *Deep residual learning for image recognition.* Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 770–778. https://doi.org/10.1109/CVPR.2016.90

[20] Hochreiter, S., & Schmidhuber, J. (1997). *Long short-term memory.* Neural Computation, 9(8), 1735–1780. https://doi.org/10.1162/neco.1997.9.8.1735

[21] Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, Ł., & Polosukhin, I. (2017). *Attention is all you need.* Advances in Neural Information Processing Systems (NeurIPS), 30. https://arxiv.org/abs/1706.03762

[22] Zhu, Q., Zhang, L., Yang, W., & Wang, J. (2020). *Application of convolutional neural networks in materials discovery and design.* npj Computational Materials, 6(1), 85. https://doi.org/10.1038/s41524-020-00354-3

[23] Lee, J., Seko, A., Shitara, K., Nakayama, K., & Tanaka, I. (2016). *Prediction model of band gap for inorganic compounds by combination of density functional theory calculations and machine learning techniques.* Physical Review B, 93(11), 115104. https://doi.org/10.1103/PhysRevB.93.115104

[24] Sanchez-Lengeling, B., & Aspuru-Guzik, A. (2018). *Inverse molecular design using machine learning: Generative models for matter engineering.* Science, 361(6400), 360–365. https://doi.org/10.1126/science.aat2663

[25] Choudhary, K., Garrity, K. F., Reid, A. C. E., DeCost, B. L., Biacchi, A. J., Hight Walker, A. R., ... & Tavazza, F. (2021). *The joint automated repository for various integrated simulations (JARVIS) for data-driven materials design.* npj Computational Materials, 7(1), 12. https://doi.org/10.1038/s41524-021-00542-2