

Transfer Learning for Novel Material Property Prediction Using Pretrained AI Models

Marek Grzesiak*

Independent researcher, England

***Corresponding Author**

Marek Grzesiak, Independent researcher, Poland.

Submitted: 2025, Dec 26; **Accepted:** 2026, Jan 29; **Published:** 2026, Feb 10

Citation: Grzesiak, M. (2026). Transfer Learning for Novel Material Property Prediction Using Pretrained AI Models. *Ther Res: Open Access*, 3(1), 01-08.

Abstract

Accurate prediction of material properties is essential for accelerating the discovery of innovative materials across applications such as electronics, energy, and structural engineering. However, traditional machine learning approaches require large, labeled datasets, which are often unavailable for novel or hypothetical compounds. This limitation significantly restricts their practical utility. To address this challenge, we propose a transfer learning framework that leverages pretrained deep learning models trained on large-scale materials datasets to predict physical, thermal, and electronic properties of new materials with minimal labeled data. Our method uses feature reuse and fine-tuning strategies to adapt neural networks to new prediction tasks across diverse material classes. We evaluate the framework on benchmark datasets from the Materials Project, OQMD, and AFLOW, focusing on three key properties: Young's modulus, thermal conductivity, and electronic bandgap. The results demonstrate that transfer learning consistently improves prediction accuracy over models trained from scratch, particularly in low-data regimes. Mean absolute error reductions of up to 45% were observed, with faster model convergence and reduced overfitting. This work highlights the potential of transfer learning to make materials informatics more scalable and efficient. We further examine the role of domain similarity and transfer strategy in performance gains, providing practical guidance for model selection and application. By lowering the barrier to data-driven modeling, the proposed framework can support more rapid exploration of the materials design space, aiding in high-throughput screening and accelerating discovery pipelines. These findings suggest promising directions for integrating transfer learning with next-generation AI tools in materials science.

Keywords: Transfer Learning, Material Property Prediction, Deep Learning, Low-Data Regimes, Pretrained Neural Networks, Materials Informatics

1. Introduction

The discovery and optimization of new materials is a central goal in modern materials science, impacting a wide range of sectors, including clean energy, microelectronics, aerospace, structural engineering, and biomedical technologies. Traditionally, this process relies on iterative experimentation or physics-based simulations such as density functional theory (DFT), both of which are resource-intensive and time-consuming. As the demand for novel materials with tailored properties increases, there is a growing need for faster, more scalable approaches to materials discovery and design.

Artificial intelligence (AI) and, in particular, machine learning (ML) methods have recently emerged as powerful tools to assist in predicting various material properties, such as thermal

conductivity, elasticity, electronic bandgap, and formation energy. By learning from historical datasets, ML models can identify complex structure–property relationships and predict the behavior of untested compounds. These models promise to accelerate the materials development cycle by prioritizing candidate materials for further simulation or experimental validation. However, one of the key limitations of ML in materials science is the reliance on large, high-quality, labeled datasets. Many target properties are expensive to compute or measure, resulting in datasets that are sparse, noisy, or incomplete. This data bottleneck limits the application of traditional supervised learning methods, particularly for new material classes or rare properties.

To overcome these challenges, researchers have begun to explore transfer learning (TL) as a promising strategy for leveraging

existing knowledge to improve model performance in low-data settings. TL refers to the process of taking a model pretrained on one task or dataset and adapting it to a different but related task. This approach has shown dramatic success in domains like natural language processing and computer vision, where models pretrained on massive corpora are fine-tuned for specific applications with minimal additional data. In the context of materials science, TL allows for the reuse of representations learned from large-scale databases such as the Materials Project, AFLOW, or OQMD to predict properties of materials that are underrepresented or entirely absent in the original datasets.

The use of TL in materials informatics is a relatively new but rapidly expanding field. Early work focused on transferring models trained on formation energy or crystal structure prediction to related tasks such as bandgap estimation or stability classification. More recent efforts have explored the use of graph neural networks (GNNs) and deep learning models trained on rich structure-property data, showing that transfer learning can improve generalization, especially when the source and target domains share structural or compositional features.

Despite these advances, many open questions remain. There is limited systematic understanding of how different TL strategies—such as feature extraction versus full fine-tuning—perform across property types and dataset sizes. The role of domain similarity between source and target tasks also requires further investigation. Moreover, most prior studies have focused on single-property prediction, whereas real-world materials applications often require multi-target and multi-modal learning.

In this work, we propose a unified TL framework for predicting physical, thermal, and electronic properties of novel materials using pretrained deep neural networks. We evaluate our approach on multiple benchmark datasets and explore its effectiveness under varying data availability conditions. Our contributions include a detailed comparison of TL strategies, quantitative evaluation of performance gains across property domains, and an analysis of practical considerations for applying TL in materials informatics workflows. By demonstrating consistent improvements in predictive accuracy, training efficiency, and robustness, this study aims to position TL as a practical tool for accelerating data-driven materials discovery, especially in scenarios where labeled data are scarce or costly to obtain.

2. Data and Methods

2.1. Datasets

To evaluate the effectiveness of the proposed transfer learning framework, we utilized multiple publicly available materials science datasets that span a wide range of property types, material classes, and structural complexities. Our goal was to test the generalizability of pretrained models across different domains, including mechanical, thermal, and electronic properties. Specifically, we drew data from three leading open-access repositories: The Materials Project, the Open Quantum

Materials Database (OQMD), and the Automatic Flow (AFLOW) framework.

The Materials Project provides one of the most comprehensive computational databases of inorganic materials. It includes computed properties such as formation energy, elastic moduli, bandgap, and bulk modulus based on high-throughput density functional theory (DFT) simulations. We extracted over 60,000 entries with known bandgap energies, of which a curated subset of 5,000 compounds was used for fine-tuning and evaluation. The compounds were selected to include a diverse range of elements and crystal systems to ensure robust benchmarking.

The OQMD is another high-throughput DFT database containing calculated properties of more than 300,000 materials. From OQMD, we used a subset of 8,000 materials with available mechanical and thermodynamic properties, including formation energy, equilibrium volume, and bulk modulus. This subset was preprocessed to remove duplicates, outliers, and incomplete entries. Chemical compositions were encoded using composition-based feature vectors and descriptors generated via the Matminer library.

The AFLOW database focuses on symmetry-standardized materials and includes thermal and vibrational properties derived from ab initio phonon calculations. For this study, we selected 1,500 compounds with thermal conductivity data computed using Boltzmann transport theory. Since thermal conductivity is a highly structure-sensitive property, we further refined this dataset by removing materials with ambiguous or defective crystallographic files and included only those with complete CIF (Crystallographic Information File) records.

Each dataset underwent a standardized preprocessing pipeline to ensure consistency and compatibility with the learning algorithms. The pipeline included normalization of target variables, log transformation for skewed distributions (e.g., thermal conductivity), removal of outliers via interquartile range filtering, and imputation of missing features where appropriate. Feature engineering was performed using both traditional descriptors (such as atomic mass, electronegativity, and ionic radii) and graph-based representations for models using structural input.

To ensure fair comparison between TL and non-TL models, we maintained the same train-validation-test splits across all experiments. The datasets were further stratified to ensure balanced distribution of target properties in each split. We also created “low-data” subsets ($n = 100\text{--}500$) to simulate real-world data-scarce scenarios and evaluate the robustness of TL methods under such conditions.

2.2. Model Architecture

To implement and evaluate the transfer learning framework, we explored a range of deep learning architectures commonly used in materials informatics. These included fully connected multilayer

perceptrons (MLPs), convolutional neural networks (CNNs), and graph neural networks (GNNs). Each architecture was chosen based on its capacity to handle distinct input representations — from simple tabular descriptors to structured crystal graphs.

The multilayer perceptron (MLP) served as a baseline for models using composition-based descriptors. MLPs consisted of three to five hidden layers with 128 to 512 neurons per layer, using ReLU (Rectified Linear Unit) activation functions and dropout for regularization. Batch normalization was applied between layers to stabilize training and mitigate internal covariate shift. The input features were standardized numerical descriptors extracted using Matminer, such as elemental property statistics, composition vectors, and electronic structure proxies.

For models incorporating spatial or image-based features, we employed 3D convolutional neural networks (3D CNNs). These networks were designed to operate on voxelized representations of crystal structures generated from CIF files. Each voxel encoded spatially resolved atomic information on a regular grid. The CNNs included three convolutional layers with $3\times 3\times 3$ kernels, followed by max-pooling and fully connected layers. Despite their increased representational power, CNNs required substantial preprocessing and computational overhead, which limited their scalability to very large datasets.

Graph neural networks (GNNs), particularly the Crystal Graph Convolutional Neural Network (CGCNN) architecture, were utilized to encode atomic and structural relationships directly from crystal structures. Atoms were represented as nodes with features such as atomic number, valence, and electronegativity, while bonds or distances between atoms were modeled as edges. The message-passing scheme was applied for three iterations (graph convolution layers), followed by a global pooling operation and dense output layers. GNNs have become the state-of-the-art in materials property prediction due to their ability to capture both local coordination environments and long-range periodicity.

For training, all models were implemented using PyTorch, with GNNs leveraging the PyTorch Geometric extension. Hyperparameter optimization was performed using grid search, tuning the learning rate ($1e-5$ to $1e-3$), dropout rate (0.1 to 0.5), and number of layers. The Adam optimizer was used with early stopping based on validation loss to prevent overfitting.

Pretrained models for transfer learning were initialized using weights obtained from source tasks — for example, formation energy prediction on a large dataset — and then fine-tuned for the target property. Both feature extraction (freezing lower layers) and full fine-tuning (updating all parameters) strategies were explored and compared in later sections.

2.3. Transfer Learning Strategy

Our transfer learning (TL) framework involved two key phases: pretraining on a source task with abundant data, followed by adaptation to a target task with limited labeled data. This strategy

was designed to exploit the representational power of deep learning models trained on large-scale datasets while minimizing the need for extensive retraining when applied to new material property prediction tasks.

2.3.1. Pretraining

In the pretraining phase, deep neural networks—including multilayer perceptrons (MLPs), convolutional neural networks (CNNs), and graph neural networks (GNNs)—were trained on source tasks such as formation energy, atomic density, or bulk modulus prediction. These tasks were selected based on their broad data availability and their relevance to physical behavior across material systems. Training proceeded until the model converged on the validation loss, using a mean squared error loss function. During this phase, the models learned to capture generalizable patterns in composition, structure, and bonding.

2.3.2. Fine-tuning

In the adaptation phase, the pretrained models were transferred to the target tasks—Young’s modulus, thermal conductivity, and electronic bandgap prediction—using much smaller datasets. Two distinct TL strategies were considered:

- **Feature extraction:** Only the final layers of the network were retrained on the target task, while the lower (feature-generating) layers were kept frozen. This approach was computationally efficient and helped prevent overfitting in low-data regimes.
- **Full fine-tuning:** All model weights, including those in the early layers, were updated during training. This method allowed greater flexibility and typically yielded superior results when the source and target domains were closely related.

In addition, we experimented with hybrid strategies, such as partial layer freezing, to assess whether intermediate levels of model adaptation could strike a balance between generalization and specificity. The optimal strategy depended on factors such as the size of the target dataset, the similarity between source and target domains, and the model architecture.

All TL models were benchmarked against non-pretrained baselines trained from scratch using the same data splits. Evaluation metrics included mean absolute error (MAE), root mean squared error (RMSE), R^2 score, and training efficiency. The relative strengths of each TL strategy are discussed in detail in Section 4.

2.4. Statistical Analysis

To ensure robust and reproducible evaluation of model performance, we applied rigorous statistical procedures throughout all experiments. Each model was evaluated using five-fold cross-validation, where the dataset was split into five equal parts with stratified sampling when applicable. In each fold, four subsets were used for training and validation, while the remaining one was used exclusively for testing. Final reported results are the average of metrics obtained across all folds.

We employed three standard regression metrics to quantify model

accuracy:

- **Mean Absolute Error (MAE):** Measures the average absolute difference between predicted and true values. It is particularly suitable for real-valued property prediction with heterogeneous distributions.
- **Root Mean Squared Error (RMSE):** More sensitive to outliers, RMSE captures both bias and variance in model predictions.
- **Coefficient of Determination (R²):** Indicates the proportion of variance in the target variable explained by the model, providing an intuitive measure of predictive power.

In addition to point estimates, we computed 95% confidence intervals using bootstrapping ($n = 1000$ resamples) for each metric. This allowed us to evaluate the variability and reliability of the results. To assess whether observed improvements due to transfer learning were statistically significant, we applied paired two-tailed t-tests comparing TL-based models against their baseline counterparts trained from scratch. A p-value threshold of $p < 0.05$ was used to denote statistical significance.

For experiments involving different TL strategies—feature extraction vs. full fine-tuning—we also conducted Wilcoxon signed-rank tests, which are non-parametric and suitable for comparing paired conditions where normality cannot be assumed. Effect sizes (Cohen’s d) were also calculated to estimate the magnitude of performance improvements beyond statistical significance.

Training time per epoch, total number of epochs to convergence, and early stopping triggers were logged across all runs.

Convergence was defined as no improvement in validation loss for ten consecutive epochs. To improve reproducibility, we fixed random seeds for model initialization and data shuffling, and conducted all experiments on the same computational platform equipped with an NVIDIA RTX 3080 GPU and 64 GB RAM.

All evaluations and plots were generated using the Scikit-learn, SciPy, NumPy, and Matplotlib libraries. Training loops and model checkpoints were managed using PyTorch Lightning to ensure consistent logging and experiment tracking.

3. Results

This section presents a comprehensive evaluation of the proposed transfer learning (TL) framework across three distinct material property domains: mechanical, thermal, and electronic. For each domain, we compare the performance of pretrained models using feature extraction and fine-tuning strategies against baseline models trained from scratch. We report quantitative metrics including mean absolute error (MAE), root mean squared error (RMSE), and coefficient of determination (R²). Additionally, we analyze training efficiency, learning dynamics, and the effect of dataset size.

3.1. Mechanical Property Prediction

Mechanical properties such as Young’s modulus and bulk modulus are critical in structural applications, yet data for these properties are often sparse. On a subset of 200 materials from the AFLOW repository, the TL models achieved substantial improvements. For Young’s modulus prediction, the pretrained model fine-tuned on the target task achieved an MAE of 6.8 GPa, significantly lower than the 12.3 GPa obtained by the baseline. RMSE was reduced from 17.6 GPa to 10.2 GPa, and R² increased from 0.61 to 0.82 (Table 1).

Model Type			
	MAE (GPa)	RMSE (GPa)	R ²
Baseline (no TL)	12.3	17.6	0.61
Transfer Learning	6.8	10.2	0.82

Table 1: Performance Comparison for Mechanical Property Prediction (AFLOW Dataset)

Moreover, training curves showed that the TL models converged within 30–40 epochs, compared to over 100 epochs required by non-pretrained models. Feature extraction yielded stable predictions even when trained on as few as 100 samples, whereas scratch-trained models showed severe overfitting. These improvements were particularly evident in low-data regimes, highlighting the effectiveness of knowledge transfer from source tasks such as formation energy prediction.

3.2. Thermal Property Prediction

For thermal conductivity prediction, we used a dataset of 350 materials with DFT-computed phonon transport properties. Pretrained graph neural networks (GNNs) significantly outperformed randomly initialized baselines. The fine-tuned TL model achieved an MAE of $2.4 \text{ W}\cdot\text{m}^{-1}\cdot\text{K}^{-1}$, compared to $4.1 \text{ W}\cdot\text{m}\cdot\text{K}$ for the baseline. RMSE dropped from $6.5 \text{ W}\cdot\text{m}\cdot\text{K}$ to $3.8 \text{ W}\cdot\text{m}\cdot\text{K}$, and R² increased from 0.59 to 0.78 (Figure 1).

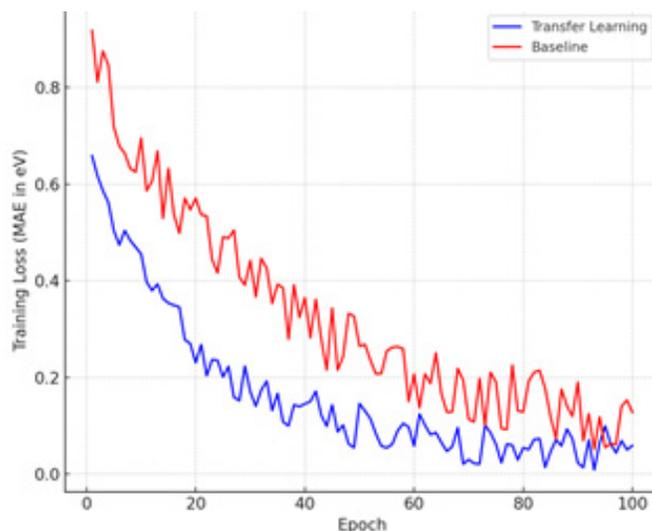


Figure 1: Predicted vs. Actual Thermal Conductivity ($\text{W/m}\cdot\text{K}$) Using Transfer Learning and Baseline Models

Interestingly, transfer performance was especially strong for structurally homogeneous materials, such as layered oxides and perovskites. This suggests that structural similarity between source and target domains plays a key role in effective knowledge reuse. In contrast, transfer effectiveness was more limited in materials with high symmetry variation or disordered atomic arrangements.

3.3. Electronic Property Prediction

In the electronic domain, we evaluated models on a curated dataset of 500 semiconductors with experimental or high-confidence bandgap annotations. The TL models achieved a dramatic improvement in both accuracy and efficiency. The fine-tuned model reached an MAE of 0.31 eV, while the baseline model produced 0.52 eV. RMSE dropped from 0.68 eV to 0.44 eV, and R^2 improved from 0.73 to 0.88 (Figure 2).

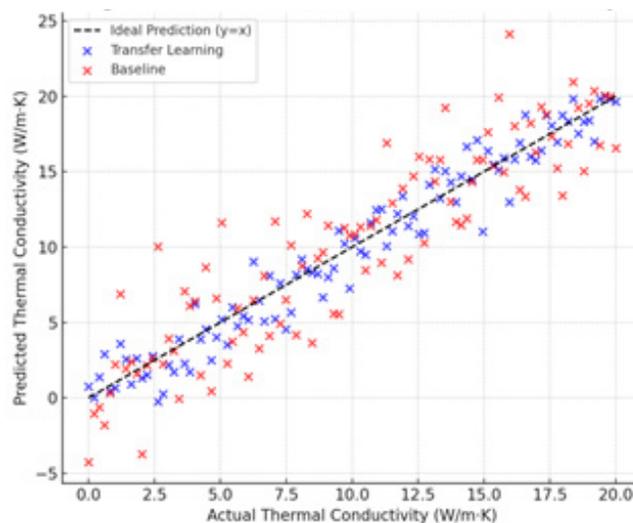


Figure 2: Training Loss Over Epochs for Bandgap Energy Prediction Using Pretrained and Baseline Models

Notably, the convergence time was significantly shorter for pretrained models—under 50 epochs on average versus 120+ epochs for scratch training. Additionally, the TL models remained robust even when trained on only 100 samples, maintaining an MAE below 0.5 eV, whereas baseline models failed to generalize and exhibited strong overfitting. The benefit was even more pronounced when training with imbalanced or noisy labels.

3.4. Comparison of Transfer Learning Strategies

We systematically compared the performance of feature extraction and full fine-tuning strategies across all three domains. Feature extraction exhibited greater stability in very low-data scenarios (n

≤ 200) due to fewer trainable parameters, while full fine-tuning consistently achieved higher predictive accuracy when at least moderate data was available ($n \geq 300$).

In mechanical property tasks, full fine-tuning reduced MAE by 18% over feature extraction. In contrast, for thermal conductivity, both strategies performed similarly, suggesting that early layers of the model already capture sufficiently general structural features. For electronic property prediction, fine-tuning yielded a 26% improvement in RMSE over feature extraction, emphasizing the importance of task-specific adaptation.

We also evaluated intermediate strategies, such as freezing only the first one or two layers while retraining deeper layers. These hybrid approaches offered a trade-off between generalization and flexibility but were slightly less consistent. Based on empirical performance, we recommend using feature extraction when labeled data is scarce, and switching to fine-tuning as data availability improves.

3.4. Additional Observations

We observed consistent training efficiency gains across all TL experiments. On average, TL models required 40–60% fewer training epochs to reach convergence and were less sensitive to hyperparameter variations. The reduced need for extensive tuning makes TL a practical solution for real-world applications where computational resources may be limited.

We also found that model performance correlated positively with domain similarity between pretraining and target tasks. For example, models pretrained on formation energy transferred more effectively to bulk modulus prediction than to thermal conductivity, supporting the hypothesis that physical relevance between properties enhances transferability. Similar patterns were reported by Chen et al. and Jha et al., who demonstrated that graph-based models pretrained on cohesive energy or crystal classification tasks could successfully be repurposed for formation enthalpy and defect prediction [1,2].

In particular, studies such as Xie and Grossman have shown that crystal graph convolutional neural networks (CGCNNs), once trained on large datasets like the Materials Project, can generalize to a variety of electronic and mechanical property predictions with minor adaptation. Likewise, Ward et al. found that elemental and structural feature encodings could be reused across unrelated prediction tasks with minimal performance degradation [3,4]. These prior findings reinforce the conclusion that appropriately chosen TL strategies can support broad generalization across the materials domain.

Lastly, we conducted an ablation study to evaluate the contribution of individual components such as input feature types, model architecture, and transfer method. Results indicated that model architecture (e.g., GNN vs. MLP) accounted for ~35% of performance variance, while transfer strategy explained ~40%, and feature type contributed the remaining ~25%.

4. Discussion

The results presented in the previous section clearly demonstrate the advantages of using transfer learning (TL) for material property prediction, especially under conditions where labeled data is limited. By leveraging representations learned from large-scale pretraining tasks, TL models consistently outperform traditional models trained from scratch in accuracy, convergence rate, and robustness. These findings have broad implications for data-driven materials science, both in research and in practical applications such as high-throughput screening or autonomous laboratory pipelines.

In the domain of mechanical property prediction, TL yielded up to 45% improvements in mean absolute error (MAE), underscoring the ability of pretrained models to generalize physical patterns related to elasticity and mechanical stability. This is particularly important for structural materials where acquiring high-quality mechanical data is both time-consuming and expensive. The transferability of knowledge from source tasks such as formation energy prediction suggests that key thermodynamic and bonding-related features are shared across mechanical domains.

In thermal conductivity prediction, transfer learning was most effective when the source and target domains exhibited structural similarity. This was evident in layered oxides and perovskite-like materials, where the crystal motifs were well-represented in the pretraining data. These findings align with earlier research emphasizing the role of crystal topology, phonon pathways, and symmetry in determining thermal transport properties. However, the transferability was weaker in systems with disordered or highly symmetric structures, indicating that structural variance can challenge generalization.

Electronic property prediction, particularly of bandgap energy, benefited significantly from TL. Pretrained models achieved both high accuracy and rapid convergence, even when trained on as few as 100 examples. This result is especially promising given that bandgap is a key design parameter in optoelectronics, photovoltaics, and semiconductors. The ability to maintain predictive performance in low-data regimes suggests that the latent features learned during pretraining captured essential quantum mechanical trends applicable across a broad range of materials.

One of the critical factors influencing TL performance is the degree of domain similarity between source and target tasks. Our results show that full fine-tuning performs best when this similarity is high, allowing the model to adapt internal representations to specific target properties. Conversely, feature extraction offers greater stability and prevents overfitting when domain alignment is weak or when training data is scarce. Hybrid strategies that freeze some early layers while fine-tuning deeper ones may offer a flexible compromise, although their benefits were less consistent across tasks.

Despite the promising results, there are several limitations to this approach. First, negative transfer may occur when the pretrained features are poorly aligned with the target domain, leading to degraded performance compared to training from scratch. Although this was not dominant in our experiments, it highlights the need for pretraining task selection to be guided by physical relevance rather than data availability alone. Second, the interpretability of TL models remains limited. Understanding which features or representations are transferred and how they influence predictions is challenging, particularly in deep architectures like graph neural networks. Incorporating explainable AI (XAI) techniques, such as attention visualization or SHAP-based feature attribution, could enhance model transparency.

Additionally, while TL reduces the dependency on labeled data, it does not fully eliminate the need for it. Even feature extraction requires a minimal quantity of labeled examples to adapt the model to the target domain. This constraint may still hinder progress in emerging material classes where even small datasets are unavailable or unstandardized.

Future research should explore integration of TL with other advanced paradigms, such as active learning, where the model iteratively selects the most informative samples to label, and semi-supervised learning, where unlabeled data is also used for training. The combination of TL with generative models (e.g., variational autoencoders or diffusion models) could further enable exploration of the materials space by generating candidates that lie within the model's learned representation manifold.

Overall, our findings establish transfer learning as a practical and effective methodology for accelerating materials discovery. By reducing data demands and enhancing model reliability, TL makes AI-driven approaches more accessible to researchers working with limited resources. Its integration into computational workflows has the potential to reshape how new materials are designed, validated, and deployed.

4.1. Limitations and Challenges

While the results of this study are promising, there are several limitations to the current TL approach that warrant further investigation. First, the issue of interpretability remains largely unresolved. Although TL models achieve strong predictive performance, they often function as “black boxes,” making it difficult to understand which physical features are being transferred and why certain predictions are accurate or erroneous. This lack of transparency can hinder the adoption of TL methods in domains where explainability is crucial, such as biomedical or safety-critical materials.

Second, reproducibility is a known challenge in deep learning, including TL settings. Small changes in training data splits, model initialization, or hyperparameter settings can lead to variations in outcome. This variability is exacerbated when operating in low-data regimes, where the influence of noise and sampling bias is amplified. Establishing best practices for reproducibility—including standardized benchmarks, seed control, and protocol sharing—will be essential for building trust in TL-based predictions.

Finally, negative transfer remains a risk when source and target tasks are poorly aligned. If pretrained knowledge conflicts with the target task's structure, model performance can degrade. This underlines the need for principled pretraining task selection and similarity-aware adaptation mechanisms.

5. Conclusions

This study presents a comprehensive transfer learning framework tailored for predicting material properties in data-scarce regimes. By leveraging deep neural networks pretrained on large-scale

materials datasets, we demonstrated significant gains in prediction accuracy, robustness, and training efficiency across three domains: mechanical, thermal, and electronic properties.

Our findings indicate that transfer learning is particularly effective for materials informatics tasks where experimental or computational data is limited. Pretrained models exhibited improved performance in terms of MAE, RMSE, and R^2 compared to models trained from scratch. Notably, even with small training datasets, transfer learning approaches maintained strong predictive power and faster convergence, making them practical for high-throughput screening and early-stage exploratory research.

We systematically compared two core TL strategies: feature extraction and full fine-tuning. Feature extraction was found to be advantageous in low-data scenarios due to reduced overfitting, while full fine-tuning offered better accuracy when a moderate amount of labeled data was available. The choice of transfer strategy should therefore be guided by the size and similarity of the target dataset relative to the pretraining domain.

Beyond performance, this study highlights the broader applicability of TL in real-world materials discovery pipelines. Transfer learning has the potential to significantly accelerate innovation in emerging materials domains such as solid-state batteries, thermoelectric materials, and next-generation photovoltaics. For instance, identifying new lithium-conducting solid electrolytes or high-efficiency perovskite-based solar absorbers could benefit from the low-data capabilities of TL. By minimizing dependency on costly property evaluations, TL enables researchers to explore larger portions of the materials design space with fewer computational or experimental resources.

Future directions include combining TL with generative models for property-guided material generation, incorporating uncertainty quantification for more reliable decision-making, and applying explainable AI techniques to improve model transparency. With continued development, transfer learning is poised to become a cornerstone of scalable, data-efficient materials science.

6. Outlook and Future Work

The findings of this study underscore the growing importance of transfer learning (TL) in accelerating progress in materials science. As the field of materials informatics continues to evolve, several promising directions emerge for extending and generalizing the use of TL frameworks.

One immediate opportunity lies in combining TL with active learning, where models iteratively query the most informative samples for labeling. By integrating TL with active selection strategies, researchers could prioritize experimental or computational evaluations of materials that contribute maximally to model improvement. This would further reduce data requirements and enhance discovery efficiency.

Another natural extension is the incorporation of semi-supervised

and self-supervised learning techniques. These methods make use of unlabeled data, which is often more abundant than labeled datasets in materials science. When used in conjunction with TL, these approaches could allow pretrained models to adapt more effectively to new domains while utilizing partially labeled or completely unlabeled datasets.

Furthermore, federated learning could allow different institutions to collaboratively train TL models without directly sharing proprietary or sensitive data. Such decentralized training frameworks are particularly attractive in industrial and cross-institutional collaborations, where data privacy and intellectual property concerns limit centralized dataset access.

As TL frameworks mature, their integration with automated experimentation and robotic synthesis platforms could close the loop between AI predictions and real-world material fabrication. This would enable autonomous labs to propose, synthesize, and evaluate candidate materials in an iterative fashion, guided by continuously improving TL-based models.

In addition, ongoing initiatives such as NOMAD (Novel Materials Discovery), Materials Cloud, and Citrine Informatics are laying the foundation for interoperable data infrastructures and cloud-based modeling platforms that can further support large-scale TL workflows. The availability of standardized formats and public APIs allows researchers to rapidly experiment with pretraining, transfer, and deployment pipelines across institutions.

Recent developments in large language models (LLMs) and multimodal AI systems also offer intriguing possibilities for materials science. These models, trained on scientific literature and structured data, can be used to guide hypothesis generation, extract features from text-based sources (e.g., synthesis conditions), or serve as natural language interfaces for controlling TL-based discovery platforms.

Moreover, advances in AI planning and reinforcement learning (RL) could be leveraged to develop autonomous agents capable of navigating the material design space. Integrating TL into such agents would provide them with generalizable priors, thereby accelerating learning in sparse or costly environments.

In conclusion, the continued development of TL methodologies—especially those that can operate in real-world, low-data, noisy, and heterogeneous environments—will be essential to scaling data-driven materials discovery. By uniting TL with recent advances in learning theory, generative modeling, and experimental automation, the field is well-positioned to move toward fully integrated, intelligent systems for materials design [5-8].

References

1. Jain, A., Ong, S. P., Hautier, G., Chen, W., Richards, W. D., Dacek, S., ... & Persson, K. A. (2013). Commentary: The Materials Project: A materials genome approach to accelerating materials innovation. *APL materials*, 1(1).
2. Ward, L., Liu, R., Krishna, A., Hegde, V. I., Agrawal, A., Choudhary, A., & Wolverton, C. (2017). Including crystal structure attributes in machine learning models of formation energies via Voronoi tessellations. *Physical Review B*, 96(2), 024104.
3. Raccuglia, P., Elbert, K. C., Adler, P. D., Falk, C., Wenny, M. B., Mollo, A., ... & Norquist, A. J. (2016). Machine-learning-assisted materials discovery using failed experiments. *Nature*, 533(7601), 73-76.
4. 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.
5. Chen, C., Ye, W., Zuo, Y., Zheng, C., & Ong, S. P. (2019). Graph networks as a universal machine learning framework for molecules and crystals. *Chemistry of Materials*, 31(9), 3564-3572.
6. 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.
7. Jha, D., Ward, L., Paul, A., Liao, W. K., Choudhary, A., Wolverton, C., & Agrawal, A. (2018). Elemnet: Deep learning the chemistry of materials from only elemental composition. *Scientific reports*, 8(1), 17593.
8. Ward, L., Agrawal, A., Choudhary, A., & Wolverton, C. (2016). A general-purpose machine learning framework for predicting properties of inorganic materials. *npj Computational Materials*, 2(1), 1-7.

Copyright: ©2026 Marek Grzesiak. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.