Computers and Artificial Intelligence；Vol.2, No.2; 2025
E-ISSN: 3006 -3728 P-ISSN：3006-371X
DOI: https://doi.org/10.70267/cai.25v2n2.5365
Published by: Zeus Press

Efficient Classification Method for Topological Quantum Materials Based on Graph Neural Networks and Persistent Homology Theory

Yuanyuan Xu*

School of Physics and Technology, Wuhan University, China

**Corresponding author: Yuanyuan Xu.*

Abstract

Topological quantum materials hold considerable promise for applications in quantum computing and spintronic devices due to their unique electronic properties. However, traditional Density Functional Theory (DFT) methods encounter difficulties in predicting these topological properties, including high computational costs and classification errors. This study proposes a machine learning framework that combines Graph Isomorphism Networks (GIN) with Atomic-Specific Persistent Homology (ASPH) to achieve efficient classification of topological materials by integrating global crystal structure features with local atomic topological descriptors. GIN is used to capture the global graph representation of periodic crystal structures, while ASPH extracts local features of atomic environments through multiscale topological analysis. These two feature sets are integrated following dimensionality reduction and subsequently classified using XGBoost. Principal Component Analysis (PCA) was employed to reduce the dimensionality of the high-dimensional ASPH feature vectors, thereby enhancing both model efficiency and accuracy. Experimental results indicate that this method performs exceptionally well in binary classification (topologically trivial/non-trivial), achieving an accuracy of 87.32%, which is significantly better than models based on a single feature. However, performance in ternary classification (trivial, semi-metal, topological insulator) declines to 75.04% due to class imbalance and feature overlap. The study validates the feasibility of combining graph neural networks with topological data analysis, providing an efficient computational framework for high-throughput screening of topological materials and offering new ideas for the application of multimodal feature fusion in materials science.

Keywords

topological quantum materials, graph isomorphism networks (GIN), atomic-specific persistent homology (ASPH), XGBoost classification

1. Introduction

Topological materials, including topological insulators and semimetals, exhibit extraordinary electronic properties arising from their unique band structures (Armitage et al., 2018). These materials are characterised by an insulating bulk state coexisting with conductive surface or edge states - a distinctive feature originating from nontrivial topological quantum states and protected by specific symmetries (Hasan & Kane, 2010). The symmetry-protected nature of these states renders these topological phases remarkably robust against various

perturbations, such as lattice defects, impurities, and external fields, making them exceptionally promising for quantum information processing and spintronic applications (Qi & Zhang, 2011).

The dissipationless edge states of topological insulators, for instance, provide an ideal platform for realising quantum bits (qubits) in quantum computing architectures. Their potential applications span multiple domains, including low-power electronic devices, spin transport systems, and the construction of superconducting qubits (Halász & Balents, 2012; Moore, 2010). Meanwhile, topological semimetals offer equally intriguing possibilities due to their linearly dispersive bands and unconventional Fermi-arc surface states. These features not only enable the exploration of novel electronic phases but also facilitate the discovery of exotic transport phenomena, thereby bridging condensed matter physics and materials science in unprecedented ways (Soumyanarayanan et al., 2016).

Despite significant advances in the study of topological materials, fundamental understanding and functional design remain formidable challenges. The discovery of topological phases critically depends on precise electronic structure calculations and their physical interpretation, while experimental characterisation of these quantum states often proves complex and time-consuming. This gap highlights the urgent need for robust theoretical and computational support.

Density Functional Theory (DFT), as a first-principles computational method, has become a fundamental tool for electronic structure research and has played a significant role in predicting topological properties, yet it still faces numerous limitations. Firstly, the computational complexity of DFT increases exponentially with system size, making studies of large-scale complex crystal structures and multi-element systems extremely time-consuming or even infeasible, thereby directly restricting the widespread application of high-throughput automated screening (Blaha et al., 2020). Secondly, many classical DFT approximate functionals (such as commonly used GGA and LDA) exhibit systematic errors in bandgap estimation, potentially leading to misjudgement of topological phases, particularly in narrow-gap semiconductors or near topological phase boundaries. Although more accurate methods (e.g., GW, rigorous treatment of spin-orbit coupling, etc.) improve precision, their extremely high computational costs make them unsuitable for large-scale screening (Bramer & Wei, 2020; Janotti & Van de Walle, 2009; Perdew et al., 1996). Furthermore, complex many-body effects, electron correlations, and temperature effects are not effectively captured by conventional DFT, further limiting its predictive capability and generalisability (Tran & Blaha, 2009). Consequently, the sole reliance on DFT for topological materials discovery creates a contradiction between computational resources and accuracy, driving the development of emerging empirical, data-driven, and machine learning approaches.

Recent advances in machine learning (ML) have provided promising solutions to these computational challenges 11241. ML methods can achieve comparable accuracy to conventional approaches while significantly reducing computational costs, often by several orders of magnitude (Giustino et al., 2020; Peano et al., 2021). This acceleration is made possible by learning structure-property relationships from feature representations rather than solving quantum mechanical equations explicitly (Jha et al., 2019; Xue et al., 2016). The growing availability of experimental and DFT-calculated structural databases further facilitates data-driven modelling (Stanev et al., 2018; Tao et al., 2021).

Notably, graph neural networks (GNNs) have emerged as particularly suitable for crystalline materials, where atomic structures are naturally represented as graphs (nodes = atoms, edges = interatomic interactions) (Wu et al., 2021; Xie & Grossman, 2018). Furthermore, GNNs can effectively integrate contextual information between nodes (atoms) through their message-passing mechanism, enabling them to adapt to multi-scale interactions in complex crystal structures (Fung et al., 2021). When combined with end-to-end training strategies, GNNs demonstrate outstanding performance in both topological property classification and prediction, establishing them as a crucial component of machine learning approaches in materials science. The Graph Isomorphism Network (GIN), a GNN variant specifically designed to discriminate between subtly different graph structures, exhibits discriminative power equivalent to the Weisfeiler-Lehman graph isomorphism test. Xu et al. (2019) demonstrated GIN's superior performance in graph classification tasks. This capability proves crucial for identifying topological materials where small structural variations may lead to distinct electronic states.

Atom-specific persistent homology (ASPH) was initially developed for investigating protein active sites and organic small molecules, where it characterises the contribution of individual atoms to topological features by tracking the formation and annihilation of topological structures such as cavities and rings across different scales (Cang & Wei, 2018; Otter et al., 2017; Xia et al., 2015). This approach has proven effective in elucidating the role of specific atoms in maintaining structural stability and defining functional regions (Bramer & Wei, 2020). Building upon this foundation, Jiang et al. (2021) demonstrated that ASPH could be successfully extended to analyse the topological properties of periodic crystalline structures, achieving accurate predictions of formation energies and related material properties.

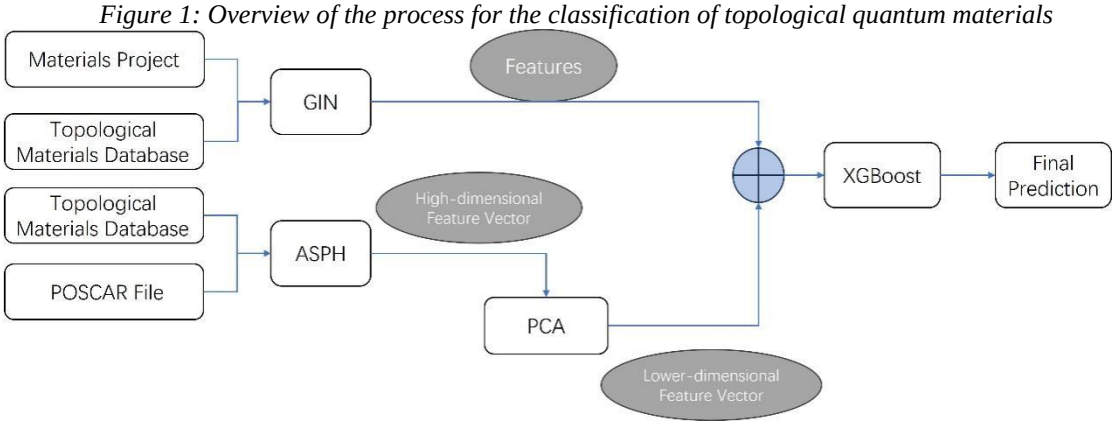
In the present study, we combine ASPH with Graph Isomorphism Networks (GIN) to construct composite feature representations for classifying the topological properties of crystalline materials. The ASPH component captures local atomic environments through its multi-scale topological descriptors, while GIN processes the global connectivity patterns of the crystal structure (de Jong et al., 2016; Raccuglia et al., 2016). This combined approach provides a more comprehensive characterisation of materials by integrating both local and structural information, thereby improving the accuracy of topological classification (Chen et al., 2019).

The successful application of this methodology not only advances our fundamental understanding of the relationship between atomic structure and topological properties but also establishes an efficient framework for high-throughput screening of topological materials. The computational efficiency of this approach, coupled with its improved predictive capability, suggests promising potential for accelerating the discovery of novel topological phases. Future research directions may include extending this method to predict other quantum topological invariants and exploring its applicability to a broader range of material systems.

2. Data Process

2.1 Overall Procedure

Figure 1: Overview of the process for the classification of topological quantum materials



The overall procedure is explained in Figure 1. The proposed processing framework consists of two main components. First, the input data for the Graph Isomorphism Network (GIN) were derived from the topological classification of materials in the Topological Materials Database (Bradlyn et al., 2017; Vergniory et al., 2019; Vergniory et al., 2022) and the crystal structures provided by the Materials Project (Jain et al., 2013). These data were then processed by the GIN to generate a subset of features for XGBoost. Second, high-dimensional feature vectors were extracted through atom-specific persistent homology (ASPH), followed by dimensionality reduction via Principal Component Analysis (PCA). Finally, the feature vectors from both components were integrated and fed into the XGBoost classifier for predictive modelling.

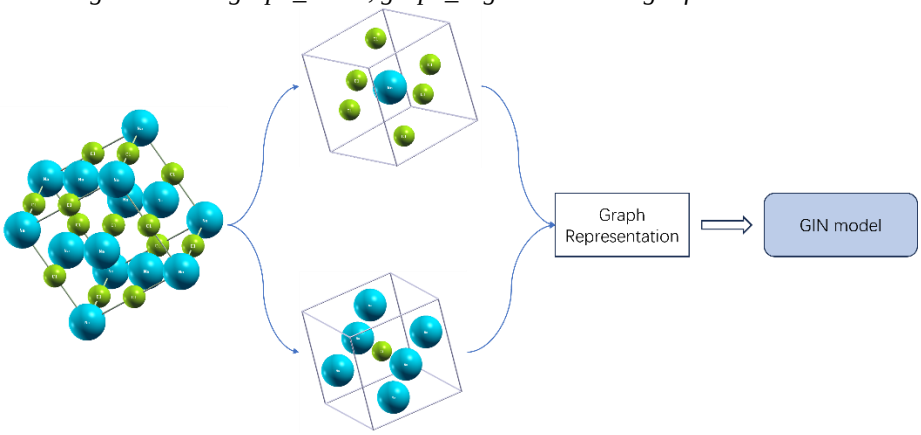
2.2 Dataset Partitioning

Following the removal of materials unavailable in the Materials Project from the Topological Materials Database, a curated dataset comprising approximately 30,000 materials was obtained. To address the dual-

classifier architecture employed in this study, the dataset was allocated equally between the two classifiers. This partitioning strategy ensures adequate data availability for each classifier's training requirements. This strategy also decouples feature generation and model training phases - specifically, the GIN-derived features for XGBoost were exclusively generated using the reserved subset, thereby preventing data leakage and enhancing the model's generalisation capability.

2.3 Crystal Graph Process

Figure 2: Illustration of how to generate the GIN input vectors. According to the CGC, we find the neighbour atoms and generate the graph_nodes, graph_edges and relating information vectors



The crystalline structure can be inherently represented as a graph, where nodes correspond to atomic positions and edges denote interatomic connectivity. However, the three-dimensional periodic nature of crystals necessitates structural preprocessing to eliminate redundant periodic replications and construct graph representations compatible with graph neural network (GNN) architectures.

In this study, crystal graph construction was implemented using a Crystal Graph Coordinator (CGC) (Yamamoto, 2019), an algorithm that captures material properties through topological connectivity patterns rather than relying on spatial parameters such as bond distances. This approach prioritises the identification of dominant chemical bonding configurations and critical nearest-neighbour interactions, thereby ensuring the constructed graphs focus on chemically meaningful features. Figure 2 illustrates our method. The critical parameters of the CGC were configured with a neighbourhood cutoff radius (α) of 1.5 and a connection tolerance threshold (β) of 0.03. The neighbourhood cutoff radius (α) was determined through empirical optimisation to achieve optimal performance, while the connection tolerance threshold (β) was adopted based on recommended values from established literature.

The present study addresses two classification tasks: distinguishing topological triviality in crystalline materials, and categorising materials into trivial, semimetal (SM), and topological insulator (TI) phases. Given that crystalline topological properties exhibit symmetry-protected robustness and potential correlations with graph homomorphism characteristics, the Graph Isomorphism Network (GIN) - a GNN variant renowned for its graph classification efficacy was adopted as the principal architecture (Xu et al., 2019).

As illustrated in Figure 3, our implementation incorporates three GIN convolutional layers with residual connections, followed by global average pooling. The baseline GIN update rule:

$$\mathbf{h}_v^{(k+1)} = MLP\left((1+\epsilon)\,\epsilon\,\mathbf{h}_v^{(k)} + \sum_{u\in\mathcal{N}(v)}\ \mathbf{h}_u^{(k)}\right) \tag{1}$$

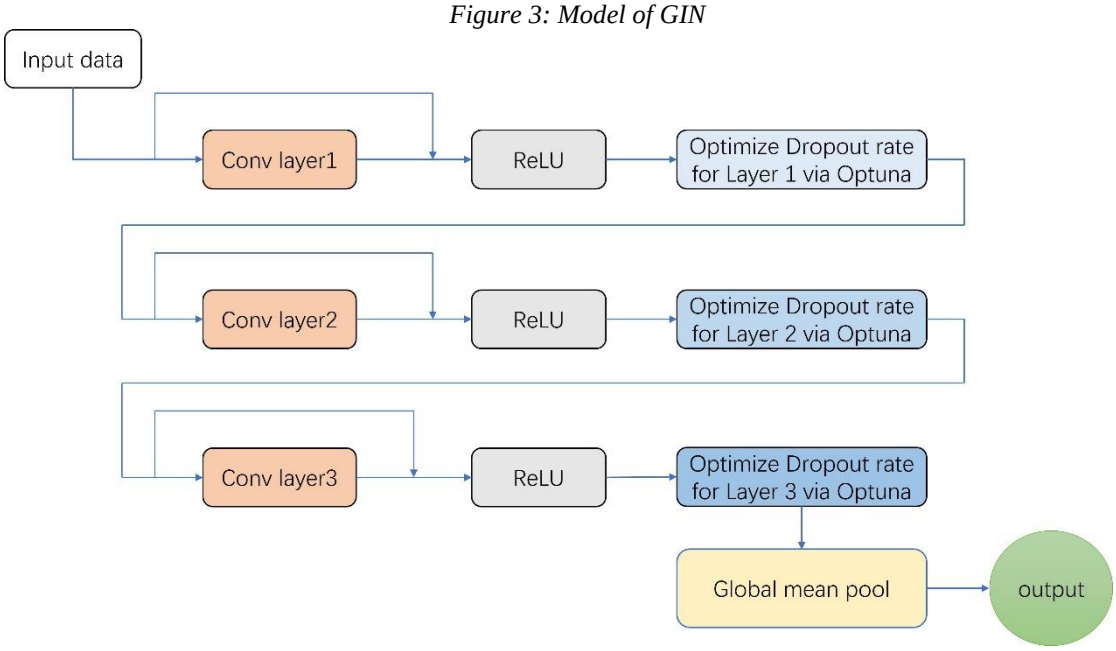
was enhanced through residual connectivity:

$$\mathbf{h}_v^{(k+1)} = MLP\left((1+\epsilon)\,\epsilon\,\mathbf{h}_v^{(k)} + \sum_{u\in\mathcal{N}(v)}\ \mathbf{h}_u^{(k)}\right) +\ \mathbf{h}_v^{k} \tag{2}$$

where ϵ (fixed at 0 in this configuration) denotes the learnable parameter, $\mathcal{N}(v)$ represents the neighbour set of node v, and MLP indicates multilayer perceptron.

Each GINConv layer was sequentially coupled with ReLU activation and layer-specific dropout regularisation, where dropout rates were optimised via the Optuna framework to establish distinct probabilities per layer. This heterogeneous dropout strategy effectively mitigates overfitting risks while preserving critical feature interactions. The architecture culminates in graph-level embedding generation through global mean pooling.

Figure 3: Model of GIN



2.4 Atom-Specific Persistent Homology (ASPH)

The acquisition of distinctive features complementary to crystal graph representations necessitates alternative analytical approaches to decode topological information embedded in crystalline structures. While conventional structural descriptors employed in GIN predominantly capture atomic species and adjacency relationships, they lack explicit encoding of atom-resolved topological signatures. This limitation motivates the adoption of Atom-Specific Persistent Homology (ASPH) as a computational tool for generating atomic-level topological characterisations (Jiang et al., 2021).

Conventional persistent homology, though effective in extracting global structural patterns, proves inadequate for periodic crystalline systems due to their complex unit cell configurations and translationally symmetric arrangements. ASPH addresses this challenge through localised topological analysis: By constructing chemical environments centred on individual atoms within the unit cell, it systematically encodes atom-specific chemical information into topological invariants through multi-scale filtration processes. This methodology generates unique topological fingerprints for individual atoms, effectively capturing both short-range bonding configurations and medium-range interaction patterns that govern material properties.

The ASPH algorithm performs atom-specific topological characterisation by constructing an enlarged cell with an 8 Å cutoff radius around each target atom, which encompasses both the central atom and its local chemical environment. For each such enlarged cell, persistent homology analysis generates Betti-number sequences (Betti-0, Betti-1, and Betti-2 barcodes) for the central atom, producing corresponding topological barcodes.

Using $BaTiO_3$ as a representative example, the feature construction process begins with computing Betti-0, Betti-1 and Betti-2 barcodes for every atom in the unit cell, followed by aggregating these barcodes according to atomic species (Ba, Ti, O). For each Betti number, five statistical descriptors are calculated:

minimum, maximum, mean, standard deviation, and sum. Specifically, Betti-0 features consider only death values (yielding 5 features), while Betti-1 and Betti-2 features incorporate birth, death, and persistence length (generating 15 features per Betti number). This results in 35 features per atomic species (5 from Betti-0 plus 15 each from Betti-1 and Betti-2).

The complete feature vector combines both local atomic contributions (105 features from 3 species) and unit cell-level statistics (additional 35 features from whole-cell barcode integration), totalling 140 non-zero features. To ensure compatibility with machine learning workflows, the feature space is standardised to 3,150 dimensions through zero-padding, accounting for all 89 naturally occurring elements in the data, plus one padding index in the 35-dimensional feature.

3. Discussion on Feature Combination

At this stage, each crystal structure possesses both ASPH-derived features and GIN classification outputs. However, significant dimensionality disparity exists between these two feature sets: the ASPH parameters comprise 3,150 dimensions (a sparse vector with numerous zero entries), while the GIN initially produces only two-dimensional logits for binary classification. To address this imbalance during feature combination, we implemented distinct dimensionality adjustment strategies for each feature type.

For the high-dimensional ASPH features, we employed Principal Component Analysis (PCA) to mitigate sparsity-related issues. The PCA retained 272 principal components, capturing 93% of the total variance while effectively eliminating noise and redundant information inherent in the original sparse matrix. This dimensionality reduction preserves the most salient topological signatures while significantly improving computational efficiency.

Regarding the GIN outputs, instead of using the final two-dimensional classification logits, we extracted the 88-dimensional pooled vector after the global mean pooling layer. This intermediate representation encodes richer structural information than the final classification output, as it preserves the global topological characteristics learned through the graph neural network's hierarchical processing. The choice of 88 dimensions reflects an optimal balance between information retention and model complexity, determined through empirical validation.

The combined feature set (272 ASPH components + 88 GIN features) provides complementary material representations: the ASPH components capture local atomic environment topology through persistent homology, while the GIN features encode global crystalline structure patterns via graph-based learning. This dual perspective enables XGBoost to leverage both atom-specific chemical information and macroscopic structural characteristics during classification. Our studies confirmed that this combined approach outperforms models using either feature set independently, demonstrating the synergistic effect of integrating geometric topology with graph neural network representations.

To elucidate the complementary nature of the combined feature vectors, we employed t-distributed Stochastic Neighbour Embedding (t-SNE) to project the high-dimensional representations onto a two-dimensional manifold. As depicted in Figure 4, distinct clustering patterns emerge between the GIN-derived features and PCA-reduced ASPH features, with the combined feature space exhibiting improved inter-class separation compared to individual modalities.

Notably, trivial and non-trivial topological classes demonstrate well-segregated distributions in the t-SNE visualisation, validating the effectiveness of our binary classification framework. However, partial overlap persists between semimetal and topological insulator clusters within the ternary classification scheme, providing geometric justification for the observed performance degradation in three-class categorisation. This spatial ambiguity in the feature manifold suggests inherent similarities in the topological descriptors of these two non-trivial phases, which may require advanced feature disentanglement techniques for enhanced discrimination.

The visualisation further corroborates our methodological rationale: the integration of GIN's global structural encoding (88D) with ASPH's local topological signatures (272D) generates a hybrid feature space that better preserves discriminative characteristics than either individual representation. This synergistic

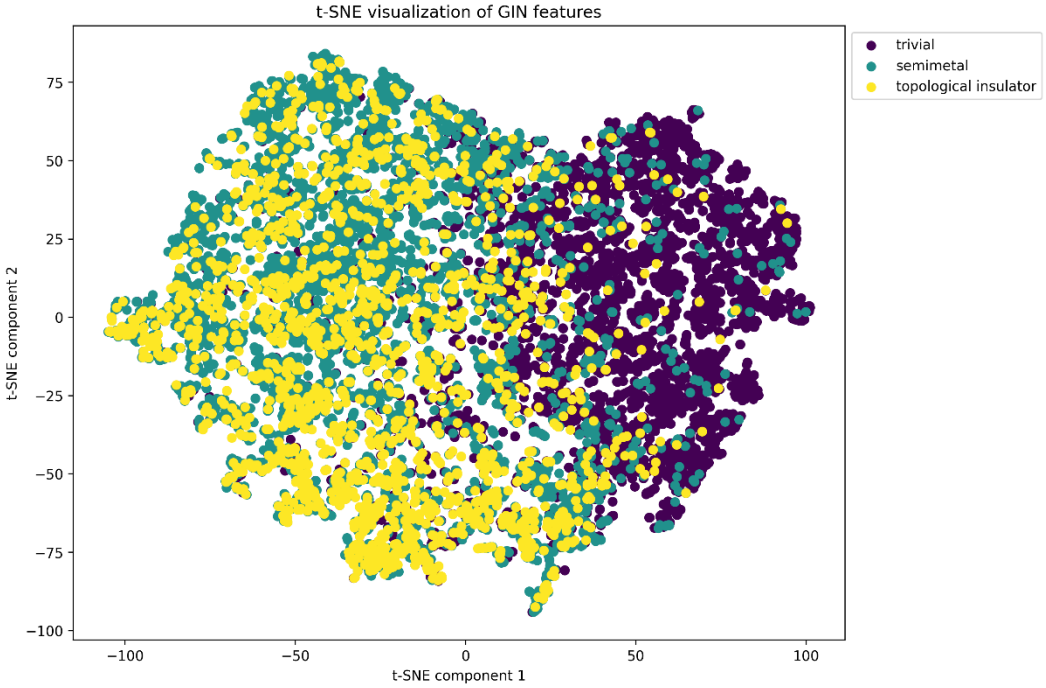
combination mitigates information loss inherent in single-modality dimensionality reduction, ultimately enhancing the downstream classifier's decision boundary optimisation.

4. Results and Discussion

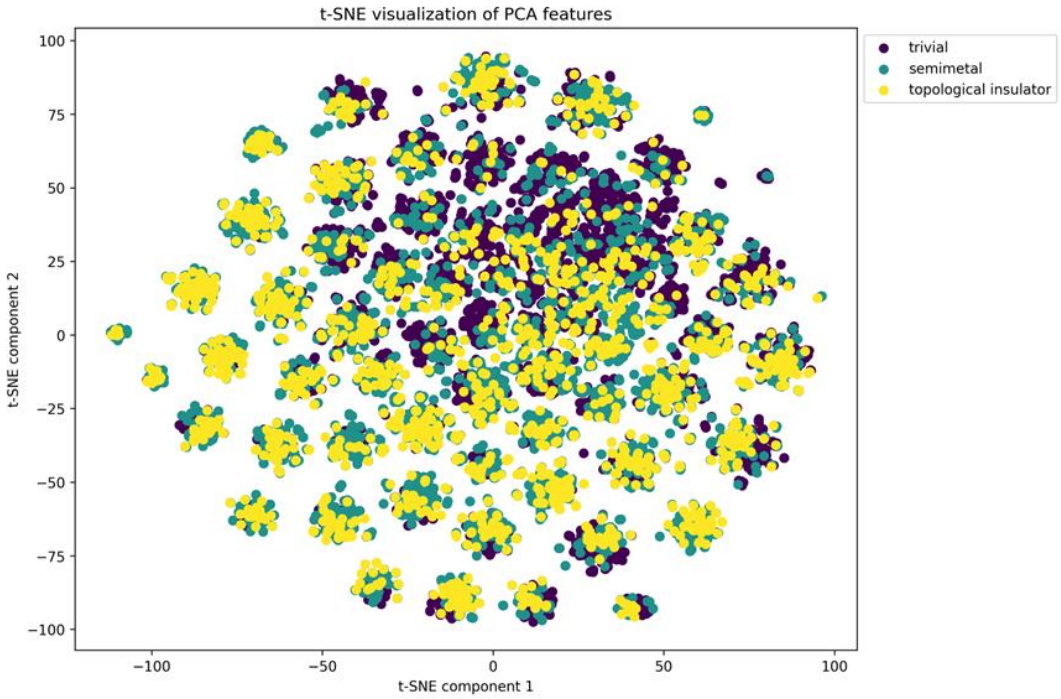
4.1 Model Validation

As evidenced in Figure 5, significant performance differences exist between models using individual feature sets (88-dimensional GIN vectors or 272-dimensional PCA-processed ASPH vectors) and our integrated approach, confirming the effectiveness of our proposed framework. This comparative analysis demonstrates that combining GIN and ASPH feature vectors provides substantially more topological information than either method alone, validating our fundamental hypothesis regarding feature complementarity.

Figure 4: Two kinds of feature vectors projected in the two-dimensional plane using t-distributed stochastic neighbour embedding. (a) GIN features, and (b) PCA features



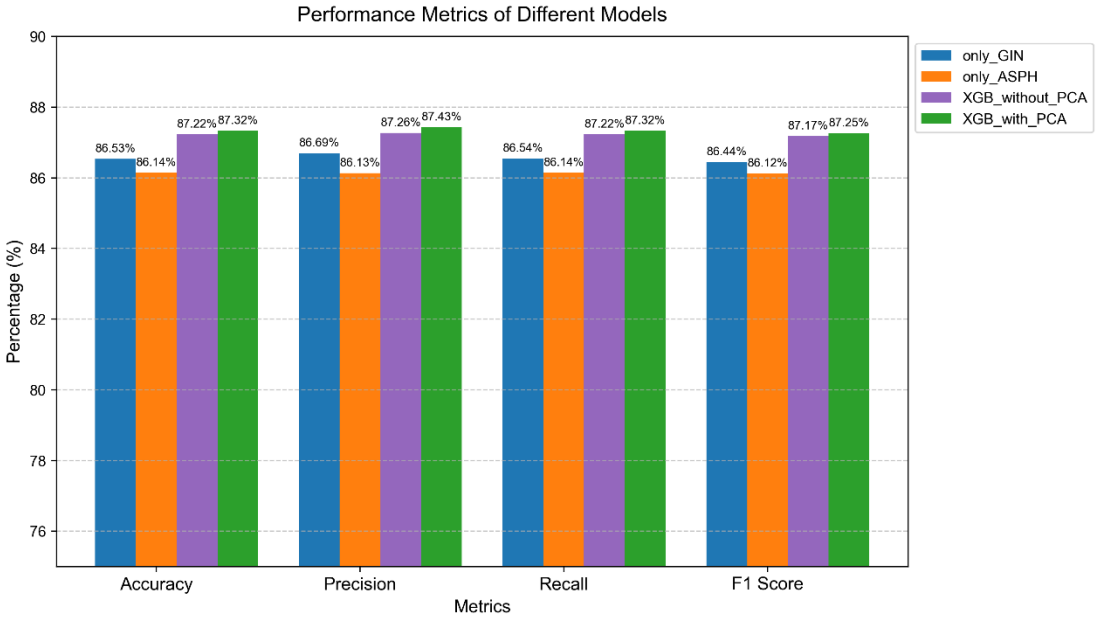
(a)



(b)

Additionally, while the model using non-PCA-reduced features achieves similar accuracy to ours, its training time is significantly longer. Given that our dataset contains tens of thousands of samples—and even larger datasets will be needed for high-throughput screening—our model significantly reduces computational costs and accelerates materials discovery.

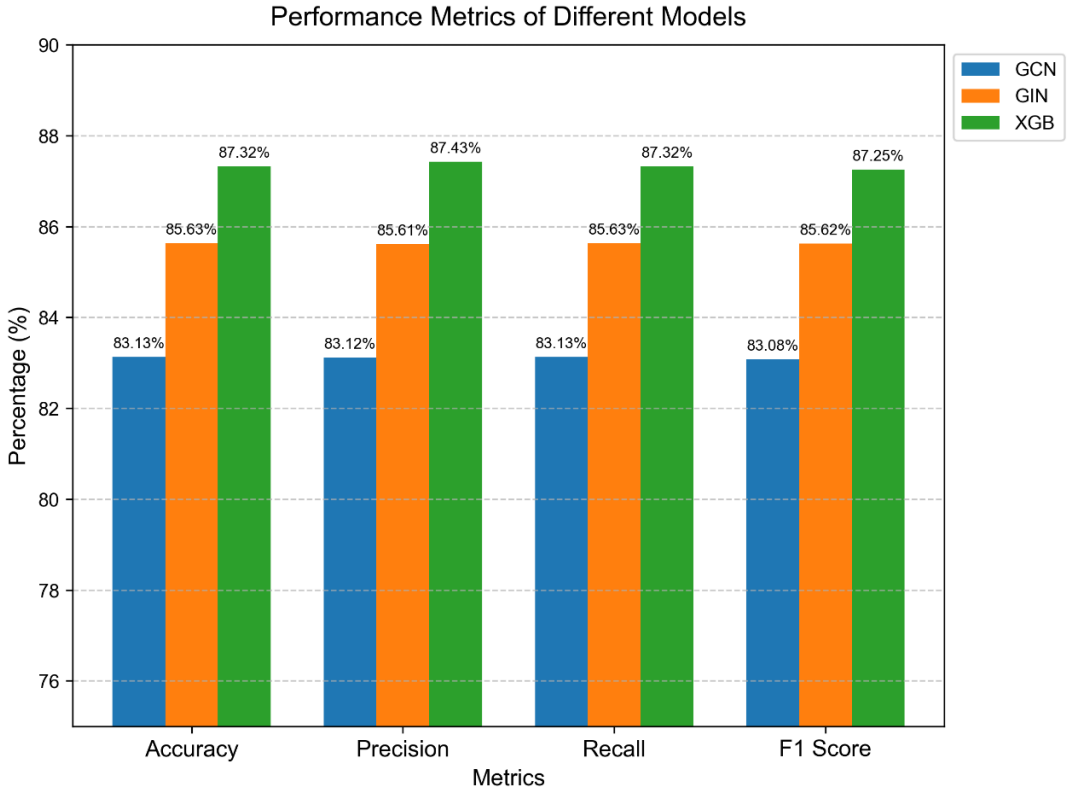
Figure 5: Performance comparison between models using only GIN parameters, only ASPH parameters, non-PCA combined parameters, and our proposed model



This optimised architecture not only accelerates individual calculations but also makes large-scale topological materials discovery computationally feasible - a crucial advancement given the exponential growth of quantum materials research. The successful balance between information preservation and dimensionality reduction represents a generally applicable strategy for other materials informatics challenges requiring multi-modal feature integration.

4.2 Classification Discussion

Figure 6: Comparison of model performance in predicting topological class in binary classification



As evidenced in Figure 6, the Graph Isomorphism Network (GIN) demonstrates superior classification performance compared to conventional graph neural networks (GNNs). We hypothesise this enhancement stems from GIN's capacity to capture features critical for crystalline topological characterisation. Specifically, the interplay between atomic connectivity patterns within unit cells and the periodic arrangements across cells - which fundamentally govern topological invariants such as Chern numbers and Z_2 indices - necessitates a model capable of discerning subtle graph isomorphism relationships. GIN's theoretical equivalence to the Weisfeiler-Lehman (WL) graph isomorphism test (Xu et al., 2019) provides a mechanistic advantage in identifying these topology-defining structural patterns.

The implemented GIN architecture employs three sequential GINConv layers with layer-wise adaptive dropout regularisation. Through systematic hyperparameter optimisation via the Optuna framework, distinct dropout rates were determined for each layer (Table 1). Notably, the initial layer exhibits near-zero dropout (0.0013), preserving critical atomic adjacency information and elemental features essential for topological discrimination. Subsequent layers demonstrate progressively increased dropout rates (0.3829 and 0.4590), effectively mitigating overfitting while maintaining hierarchical feature extraction capabilities. The architectural design aligns with the hierarchical nature of crystalline topology, where local atomic configurations establish baseline properties, while global symmetry considerations refine classification boundaries.

Table 1: Dropout rate in three GINConv layers

dropout_rate1	dropout_rate2	dropout_rate3
0.0013	0.3829	0.4590

The XGBoost classifier was trained using the combined feature vectors described previously, with hyperparameters including learning rate, max_depth, subsample, and min_child_weight optimised via the Optuna framework. During iterative optimisation, the learning rate stabilised around 0.007, while max_depth consistently converged to 7, subsample approached 1.0, and min_child_weight remained fixed at 2. This parameter convergence pattern reflects the Bayesian optimisation mechanism of Optuna, which

progressively narrows the search space toward regions yielding superior model performance. The observed parameter stability suggests that the dataset retains inherent complexity despite prior dimensionality reduction through PCA, necessitating precise configuration to balance model capacity and generalisation.

Notably, the XGBoost model trained on the combined features outperformed the standalone GIN classifier, demonstrating that the PCA-processed ASPH features introduce discriminative dimensions complementary to the graph-derived representations. For comprehensive evaluation, alternative classifiers, including neural networks and random forests, were implemented. The neural network achieved an accuracy of 84.30%, while the random forest attained 86.71%, both underperforming compared to XGBoost's superior classification accuracy. This comparative outcome highlights XGBoost's enhanced capability to leverage the hybrid feature space, effectively integrating local topological invariants from ASPH with global structural patterns encoded by GIN, thereby establishing an optimised decision boundary for crystalline topology classification.

The ternary classification experiment yielded an overall accuracy of 75.04% using XGBoost, with detailed performance metrics provided in Table 2. This performance degradation compared to binary classification primarily stems from the model's limited discriminative capability between semimetal and topological insulator phases. As shown in panels (a) and (b) of Figure 4, which separately visualise the t-SNE projections of GIN embeddings and PCA-reduced ASPH features, partial class overlap persists between semimetals and topological insulators in both feature spaces. The classification challenge is further exacerbated by inherent data limitations: topological insulator instances constitute only 14.70% of all samples, creating a pronounced class imbalance. This scarcity of representative examples hinders the model's capacity to learn distinctive boundary characteristics. While the combined feature strategy successfully improved trivial/non-trivial separation over single-modality approaches, its efficacy diminishes when addressing finer-grained categorisation within non-trivial phases, indicating the need for enhanced feature engineering.

Table 2: Performance metrics in ternary classification

	Precision	Recall	F1-Score	Support
Trivial	0.8474	0.8319	0.8396	928
Semimetal	0.7216	0.7869	0.7528	807
Topological insulator	0.4918	0.4000	0.4412	300

4.3 Comparison with Previous Work

Table 3: Comparative analysis with existing binary classification approaches

	Precision	Recall	F1-Score	Accuracy
XGBoost	0.895	0.875	0.885	0.924
Topogivity	0.856	0.741	0.795	0.872
Proposed Model	0.895	0.876	0.885	0.914
Our model	0.874	0.873	0.872	0.873

The first two rows of data are derived from the work of He et al. (2025), where the authors employed DFT calculations to generate physical property descriptors based on material characteristics. As shown in Table 3, while utilising DFT-computed data can yield better model performance, it inevitably requires significantly greater computational resources. Our work shares conceptual similarities with the approach by Rasul et al. (2024), but differs in methodology. They used Ishayev's method (Qi & Zhang, 2011) for crystal graph construction, employing an 80-20 train-test split and deep neural networks (DNNs) as the final classifier. In our study, we evenly split the dataset, using one half to train the GIN model and the other half for XGBoost training.

Although both approaches demonstrate that constructing new effective features and using multiple classifiers can improve performance over single classifiers, the differing data processing methods yield different outcomes, likely due to differences in how the features were processed and utilised. Nevertheless, both studies confirm that combining feature engineering with ensemble methods accelerates the discovery of topological properties in crystalline materials.

5. Conclusion

This study proposes an innovative machine learning framework by integrating Graph Isomorphism Networks (GIN) and Atom-Specific Persistent Homology (ASPH) for the efficient classification of topological quantum materials. GIN captures global graph representations of periodic crystal structures to extract atomic connectivity patterns and macroscopic symmetry features. In parallel, ASPH analyses fine-grained topological fingerprints of local atomic environments through multi-scale topological analysis, such as short-range bonding configurations and medium-range interaction patterns. The two types of features are fused after dimensionality reduction and fed into an XGBoost classifier for final prediction.

Experimental results show that the framework achieves 87.32% accuracy in binary classification (trivial/non-trivial), significantly outperforming single-feature models and demonstrating the synergistic effect of global-local feature fusion. However, the accuracy drops to 75.04% in ternary classification (trivial/semimetal/topological insulator), primarily due to feature overlap between semimetals and topological insulators and class imbalance issues (topological insulator samples account for only 14.7%). t-SNE visualisation further reveals that the hybrid feature space, combining GIN's 88-dimensional global structural encodings with ASPH's 272-dimensional local topological principal components, significantly enhances inter-class separation, particularly showing distinct clustering in binary classification, though geometric similarities between non-trivial phases still cause partial overlap, limiting fine-grained classification performance.

Furthermore, comparative experiments demonstrate that XGBoost outperforms neural networks and random forests in integrating both feature types, with optimised parameters (learning rate=0.007, max_depth=7) determined through Bayesian optimisation via the Optuna framework, validating the effectiveness of the feature fusion strategy. Compared to existing studies (e.g., Rasul et al.'s method), this research employs uniform dataset partitioning and heterogeneous feature engineering strategies, preventing data leakage while improving model generalisation capability, further confirming the potential of multimodal feature fusion in materials science.

Although the current method has limitations in distinguishing non-trivial phases, it establishes an efficient computational framework for high-throughput screening of topological materials by combining graph neural networks with topological data analysis, laying the foundation for predicting quantum topological invariants and cross-material system applications. Future research could further improve complex phase classification performance by introducing dynamic feature weighting, enhanced class-balancing strategies, or developing more refined topological descriptors, while extending applications to correlated quantum systems, such as high-temperature superconductors and topological superconductors.

In summary, this work bridges graph-based structural analysis with persistent homology, advancing the computational characterisation of quantum materials. While challenges remain in resolving subtle phase distinctions, the methodology lays the groundwork for future innovations in data-driven materials science, underscoring the transformative potential of multi-modal feature engineering and ensemble learning in uncovering complex material properties.

References

Armitage, N. P., Mele, E. J., & Vishwanath, A. (2018). Weyl and dirac semimetals in three-dimensional solids. *Reviews of Modern Physics*, *90*(1), Article 015001. https://doi.org/10.1103/REVMODPHYS.90.015001

Blaha, P., Schwarz, K., Tran, F., Laskowski, R., Madsen, G. K. H., & Marks, L. D. (2020). WIEN2k: An APW+lo program for calculating the properties of solids. *The Journal of Chemical Physics*, *152*(7), Article 074101. https://doi.org/10.1063/1.5143061

Bradlyn, B., Elcoro, L., Cano, J., Vergniory, M. G., Wang, Z., Felser, C., Aroyo, M. I., & Bernevig, B. A. (2017). Topological quantum chemistry. *Nature*, *547*(7663), 298-305. https://doi.org/10.1038/NATURE23268

Bramer, D., & Wei, G.-W. (2020). Atom-specific persistent homology and its application to protein flexibility analysis. *Computational and Mathematical Biophysics*, *8*(1), 1-35. https://doi.org/10.1515/CMB-2020-0001

Cang, Z., & Wei, G. W. (2018). Integration of element specific persistent homology and machine learning for protein-ligand binding affinity prediction. *International Journal for Numerical Methods in Biomedical Engineering*, *34*(2), Article e2914. https://doi.org/10.1002/CNM.2914

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. https://doi.org/10.1021/ACS.CHEMMATER.9B01294

de Jong, M., Chen, W., Notestine, R., Persson, K., Ceder, G., Jain, A., Asta, M., & Gamst, A. (2016). A statistical learning framework for materials science: Application to elastic moduli of k-nary inorganic polycrystalline compounds. *Scientific Reports*, *6*(1), Article 34256. https://doi.org/10.1038/SREP34256

Fung, V., Zhang, J., Juarez, E., & Sumpter, B. G. (2021). Benchmarking graph neural networks for materials chemistry. *npj Computational Materials*, *7*(1), Article 84. https://doi.org/10.1038/S41524-021-00554-0

Giustino, F., Lee, J. H., Trier, F., Bibes, M., Winter, S. M., Valentí, R., Son, Y.-W., Taillefer, L., Heil, C., Figueroa, A. I., Plaçais, B., Wu, Q., Yazyev, O. V., Bakkers, E. P. A. M., Nygård, J., Forn-Díaz, P., De Franceschi, S., McIver, J. W., Torres, L. E. F. F., Low, T., Kumar, A., Galceran, R., Valenzuela, S. O., Costache, M. V., Manchon, A., Kim, E.-A., Schleder, G. R., Fazzio, A., & Roche, S. (2020). The 2021 quantum materials roadmap. *Journal of Physics: Materials*, *3*(4), Article 042006. https://doi.org/10.1088/2515-7639/ABB74E

Halász, G. B., & Balents, L. (2012). Time-reversal invariant realization of the weyl semimetal phase. *Physical Review B*, *85*(3), Article 035103. https://doi.org/10.1103/PHYSREVB.85.035103

Hasan, M. Z., & Kane, C. L. (2010). Colloquium: Topological insulators. *Reviews of Modern Physics*, *82*(4), 3045-3067. https://doi.org/10.1103/REVMODPHYS.82.3045

He, Y., De Breuck, P.-P., Weng, H., Giantomassi, M., & Rignanese, G.-M. (2025). Machine learning on multiple topological materials datasets. *npj Computational Materials*, *11*(1), Article 181. https://doi.org/10.1038/S41524-025-01687-2

Jain, A., Ong, S. P., Hautier, G., Chen, W., Richards, W. D., Dacek, S., Cholia, S., Gunter, D., Skinner, D., Ceder, G., & Persson, K. A. (2013). Commentary: The materials project: A materials genome approach to accelerating materials innovation. *APL Materials*, *1*(1), Article 011002. https://doi.org/10.1063/1.4812323

Janotti, A., & Van de Walle, C. G. (2009). Fundamentals of zinc oxide as a semiconductor. *Reports on Progress in Physics*, *72*(12), Article 126501. https://doi.org/10.1088/0034-4885/72/12/126501

Jha, D., Choudhary, K., Tavazza, F., Liao, W.-k., Choudhary, A., Campbell, C., & Agrawal, A. (2019). Enhancing materials property prediction by leveraging computational and experimental data using deep transfer learning. *Nature Communications*, *10*(1), Article 5316. https://doi.org/10.1038/S41467-019-13297-W

Jiang, Y., Chen, D., Chen, X., Li, T., Wei, G.-W., & Pan, F. (2021). Topological representations of crystalline compounds for the machine-learning prediction of materials properties. *npj Computational Materials*, *7*(1), 28-28. https://doi.org/10.1038/S41524-021-00493-W

Moore, J. E. (2010). The birth of topological insulators. *Nature*, *464*(7286), 194-198. https://doi.org/10.1038/NATURE08916

Otter, N., Porter, M. A., Tillmann, U., Grindrod, P., & Harrington, H. A. (2017). A roadmap for the computation of persistent homology. *EPJ Data Science*, *6*(1), Article 17. https://doi.org/10.1140/EPJDS/S13688-017-0109-5

Peano, V., Sapper, F., & Marquardt, F. (2021). Rapid exploration of topological band structures using deep learning. *Physical Review X*, *11*(2), Article 021052. https://doi.org/10.1103/PHYSREVX.11.021052

Perdew, J. P., Burke, K., & Ernzerhof, M. (1996). Generalized gradient approximation made simple. *Physical Review Letters, 77*(18), 3865-3868. https://doi.org/10.1103/PHYSREVLETT.77.3865

Qi, X. L., & Zhang, S. C. (2011). Topological insulators and superconductors. *Reviews of Modern Physics, 83*(4), 1057-1110. https://doi.org/10.1103/REVMODPHYS.83.1057

Raccuglia, P., Elbert, K. C., Adler, P. D. F., Falk, C., Wenny, M. B., Mollo, A., Zeller, M., Friedler, S. A., Schrier, J., & Norquist, A. J. (2016). Machine-learning-assisted materials discovery using failed experiments. *Nature, 533*(7601), 73-76. https://doi.org/10.1038/NATURE17439

Rasul, A., Hossain, M. S., Dastider, A. G., Roy, H., Hasan, M. Z., & Khosru, Q. D. M. (2024). A machine learning based classifier for topological quantum materials. *Scientific Reports, 14*(1), 31564-31564. https://doi.org/10.1038/S41598-024-68920-8

Soumyanarayanan, A., Reyren, N., Fert, A., & Panagopoulos, C. (2016). Emergent phenomena induced by spin-orbit coupling at surfaces and interfaces. *Nature, 539*(7630), 509-517. https://doi.org/10.1038/NATURE19820

Stanev, V., Oses, C., Kusne, A. G., Rodriguez, E., Paglione, J., Curtarolo, S., & Takeuchi, I. (2018). Machine learning modeling of superconducting critical temperature. *npj Computational Materials, 4*(1), Article 29. https://doi.org/10.1038/S41524-018-0085-8

Tao, Q., Xu, P., Li, M., & Lu, W. (2021). Machine learning for perovskite materials design and discovery. *npj Computational Materials, 7*(1), Article 23. https://doi.org/10.1038/S41524-021-00495-8

Tran, F., & Blaha, P. (2009). Accurate band gaps of semiconductors and insulators with a semilocal exchange-correlation potential. *Physical Review Letters, 102*(22), Article 226401. https://doi.org/10.1103/PHYSREVLETT.102.226401

Vergniory, M. G., Elcoro, L., Felser, C., Regnault, N., Bernevig, B. A., & Wang, Z. (2019). A complete catalogue of high-quality topological materials. *Nature, 566*(7745), 480-485. https://doi.org/10.1038/S41586-019-0954-4

Vergniory, M. G., Wieder, B. J., Elcoro, L., Parkin, S. S. P., Felser, C., Bernevig, B. A., & Regnault, N. (2022). All topological bands of all nonmagnetic stoichiometric materials. *Science, 376*(6595), Article eabg9094. https://doi.org/10.1126/SCIENCE.ABG9094

Wu, Z., Pan, S., Chen, F., Long, G., Zhang, C., & Yu, P. S. (2021). A comprehensive survey on graph neural networks. *IEEE Transactions on Neural Networks and Learning Systems, 32*(1), 4-24. https://doi.org/10.1109/TNNLS.2020.2978386

Xia, K., Feng, X., Tong, Y., & Wei, G. W. (2015). Persistent homology for the quantitative prediction of fullerene stability. *Journal of Computational Chemistry, 36*(6), 408-422. https://doi.org/10.1002/JCC.23816

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), Article 145301. https://doi.org/10.1103/PHYSREVLETT.120.145301

Xu, K., Jegelka, S., Hu, W., & Leskovec, J. (2019). How powerful are graph neural networks? *arXiv preprint*, arXiv:1810.00826. https://doi.org/10.48550/arXiv.1810.00826

Xue, D., Balachandran, P. V., Hogden, J., Theiler, J., Xue, D., & Lookman, T. (2016). Accelerated search for materials with targeted properties by adaptive design. *Nature Communications, 7*(1), Article 11241. https://doi.org/10.1038/NCOMMS11241

Yamamoto, T. (2019). *Crystal graph neural networks for data mining in materials science*. Research Institute for Mathematical and Computational Sciences. https://www.researchgate.net/profile/Takenori-Yamamoto-2/publication/333667001_Crystal_Graph_Neural_Networks_for_Data_Mining_in_Materials_Science/links/5d008fb8a6fdccd13094038c/Crystal-Graph-Neural-Networks-for-Data-Mining-in-Materials-Science.pdf

Funding

This research received no external funding.

Conflicts of Interest

The authors declare no conflict of interest.

Acknowledgment

This paper is an research about: *Efficient Classification Method for Topological Quantum Materials Based on Graph Neural Networks and Persistent Homology Theory*. I sincerely appreciate the support received during this research. Special thanks to Ashiqur Rasul for providing part of essential data and valuable discussions on datasets. I also thank the anonymous reviewers for their constructive suggestions, all comments from the anonymous reviewers greatly improved this manuscript.