

Prediction of Enantioselectivity: A Comparative Study of Chiral-Aware Neural Network Architectures

Sanshray Pradeep

Received December 9, 2024

Accepted December 20, 2025

Electronic access January 15, 2026

Chirality plays a central role in asymmetric catalysis and drug development, yet accurately predicting enantioselectivity remains a significant challenge due to the interplay of steric, electronic, and conformational effects. This work presents a structured literature review evaluating classical machine-learning (ML) approaches and recent chiral-aware graph neural networks (GNNs) for stereochemistry-dependent prediction tasks. A systematic search identified 163 publications, of which 14 met predefined inclusion criteria requiring methodological clarity, quantitative benchmarks, and stereochemical relevance. This review addresses a gap in synthesis: although several chiral-aware architectures have been proposed and individually benchmarked, prior studies use different datasets, metrics, and task setting, making direct comparison difficult. Across six benchmark families spanning local stereocenter tasks (R/S configuration, optical rotation, enantiomer ranking) and global chirality tasks (HTS QSAR, enantioselective catalysis modeling, and general chiral property prediction), chiral-aware GNNs such as ChIRo, ChiENN, Tetra-DMPNN, MolKGNN, and ChiPGT consistently outperform classical descriptor-based models and non-chiral GNNs. These advantages arise from torsion-aware embeddings, order-sensitive message passing, and geometry-aware latent representations. While performance depends on dataset quality and stereochemical diversity, the synthesis highlights the limitations of traditional models and underscores the emerging role of chiral-aware neural architectures in accelerating asymmetric reaction design and stereoselective drug discovery.

Introduction

Small organic molecules form the basis of most modern pharmaceuticals, agrochemicals, and materials. However, the discovery and optimization of these molecules remains a costly and time-intensive process, with the development of a single approved drug requiring an average of 14 years and approximately \$2 billion in investment¹. One of the most important—and most challenging—factors influencing small-molecule drug design is chirality. Many molecules exist as *enantiomers*, which share the same connectivity but differ in their three-dimensional arrangement. These mirror-image structures can exhibit drastically different biological activities, as exemplified by the thalidomide tragedy, where the (R)-enantiomer acted as a sedative while the (S)-enantiomer caused severe birth defects^{2,3}. As a result, modern drug development increasingly demands highly enantioselective synthesis and predictive tools for ensuring the correct stereoisomeric form.

Traditional approaches to modeling and predicting enantioselectivity—such as linear free-energy relationships, descriptor-based quantitative structure–activity relationships (QSAR), and density functional theory (DFT)—often require handcrafted features, significant expert intervention, or prohibitive computational cost. Recent advances in neural

networks, particularly graph neural networks (GNNs), provide an alternative by learning molecular features directly from structure. However, early GNNs struggled to distinguish enantiomers because conventional message-passing operations are inherently insensitive to 3D chirality. This limitation led to the development of a new class of chiral-aware neural networks, such as Tetra-DMPNN, ChIRo, MolKGNN, EnzyKR, and ChiENN, which incorporate stereochemical information into their representations.

Although these models have been introduced in separate studies, a comprehensive comparison of their mechanisms, performance, and limitations has not been thoroughly presented at the high school publication level. A systematic review is therefore essential to clarify the state of the field, highlight patterns across benchmarks, and identify emerging opportunities. This work is a structured literature review rather than an experimental study, synthesizing current advances in machine-learning methods for predicting enantioselectivity.

To guide this review, we explicitly pose the following research question:

How do novel chiral-aware graph neural network architectures and classical machine-learning models compare in their ability to represent stereochemistry and predict enantioselective outcomes?

Methodology for the Literature Review

To ensure a comprehensive and unbiased survey of stereochemistry-aware predictive models, we adopted a structured literature review methodology. Searches were conducted using Google Scholar, PubMed, arXiv, and Semantic Scholar with keywords including “enantioselectivity prediction,” “chirality GNN,” “torsion-aware models,” “asymmetric catalysis machine learning,” and “stereochemical prediction.” A total of 163 publications were initially retrieved. After screening titles and abstracts, 43 articles remained. Full-text evaluation was conducted using three inclusion criteria:

1. the study introduced or evaluated ML/GNN methods for stereochemical or enantioselective tasks,
2. quantitative performance benchmarks were reported, and
3. the stereochemical representation method was clearly described.

Fourteen papers met these criteria and were included in the final review.

Literature Review

Initial efforts in enantioselectivity prediction were grounded in regression-based models, relying heavily on molecular descriptor engineering. Researchers like Eric Jacobsen and Matthew Sigman were pioneers in integrating quantitative structure-activity relationships (QSAR) and steric/electronic parameters into predictive models. Sigman’s application of Sterimol parameters, which quantify steric effects, is particularly noteworthy^{4,5}. By correlating these parameters with reaction outcomes, Sigman’s team established linear free energy relationships that provided foundational insights into asymmetric catalysis.

As data availability and computational power grew, ML methods like random forests and support vector machines became prevalent. Piras et al. leveraged random forests to predict enantioselectivity in chiral stationary phases, emphasizing the importance of balanced datasets and aggregation strategies⁶. These methods offered improved accuracy compared to traditional regression models but still relied on manual feature extraction.

Deep learning introduced a paradigm shift, allowing for end-to-end learning directly from molecular structures. Graph neural networks (GNNs) emerged as a powerful tool, but their initial implementations struggled to account for chirality. Researchers addressed this limitation through innovative Message Passing Neural Networks (MPNNs)⁷. Initial research using 2D GNNs Directed Message Passing Neural Networks

(DMPNN) with bond-level message parsing⁸, enriched embeddings with synthetic distances and angles⁹, and chiral-aware aggregation functions¹⁰ showed modest improvements over a baseline sum aggregator, highlighting opportunities for further enhancements.

Adams et al.¹¹ introduced the Chiral InterRoto-Invariant Neural Network (ChIRO), an SE(3)-invariant model designed to process torsion angles of 3D molecular conformers. By capturing molecular flexibility and respecting internal symmetry, ChIRO set a new benchmark for enantioselectivity modeling.

Similarly, Gainski et al.¹² developed the Chiral Edge Neural Network (ChiENN), a GNN variant sensitive to the spatial arrangement of atoms around a stereocenter. ChiENN distinguishes enantiomers through an order-sensitive message-passing scheme, significantly improving predictions in chiral-sensitive classifications and ranking tasks.

Liu et al.¹³ developed a GNN model named MolKGNN that addressed the interpretability gaps and enhanced the message-passing framework by introducing molecular convolution kernels to learn local chiral patterns. This design improved activity prediction in high-throughput screening (HTS) assays by identifying stereochemically relevant molecular motifs through learned kernel functions.

In enzymatic reactions, the integration of chirality-aware models like EnzyKR further advanced the field. By employing a GNN architecture and multi-task learning, EnzyKR explicitly modeled substrate chirality, addressing gaps in previous models¹⁴. This innovation was particularly effective in predicting enantioselective outcomes of hydrolase-catalyzed reactions and demonstrated the benefits of jointly modeling conversion and enantiomeric excess.

As chiral-aware architectures emerged, comparative evaluation increasingly included 3D equivariant baselines that capture molecular geometry: SchNet¹⁵, DimeNet++¹⁶, and SphereNet¹⁷. These SE(3)-equivariant models provide strong geometric priors by representing bond distances, angles, and directional information. However, they treat enantiomers as identical graphs, because their message-passing operations are permutation-invariant and collapse mirror structures into the same latent embedding. As a result, they serve as strong non-chiral 3D baselines in later stereochemistry benchmarks but lack explicit chirality-awareness.

More recently, chiral modeling has shifted toward pre-trained architectures that learn stereochemical features directly from large molecular datasets. ChiPGT¹⁸ uses a graph transformer pretrained to detect molecular handedness, then applies these representations to stereochemistry-dependent tasks with little fine-tuning. Unlike task-specific designs in ChIRO or ChiENN, ChiPGT learns a unified latent space capturing both local stereocenters and global conformational effects. This approach has demonstrated competitive performance across R/S configuration prediction, enantiomer rank-

ing, and chiral QSAR tasks, particularly in low-data scenarios.

In addition, researchers such as Connor Coley have significantly influenced the use of AI in reaction prediction. His work in deep learning for chemical synthesis has broader implications for enantioselectivity. Kristy A. Boal has also contributed by integrating cheminformatics approaches to predict stereochemical outcomes, enriching the modeling process.

The transition from feature-engineered models to chirality-aware neural networks highlights the importance of AI in enhancing enantioselectivity prediction. Models like ChiENN, ChiRo, MolKGNN, EnzyKR and ChiPGT represent state-of-the-art advancements, providing tools to optimize catalysts and design asymmetric reactions with greater predictive accuracy and broader generalizability.

Machine Learning and GNN Methods

Classical machine-learning models, while effective in specific domains, face limitations in their extensibility across different chemical reaction spaces. These models heavily rely on hand-crafted features designed around specific reaction types, and when they encounter novel chemical motifs, they often fail to capture important steric or electronic interactions that were not reflected in the original descriptor set.

Graph Neural Networks (GNNs) offer significant advantages in extensibility and adaptability because they learn molecular representations directly from structure. Their ability to encode molecular topology without predefined features allows them to generalize more effectively across chemical spaces.

However, it is not accurate to assume that GNNs always outperform classical ML models. A key methodological issue in molecular property prediction is that reported model performance frequently depends more on the evaluation protocol than on the model architecture itself. Prior studies demonstrate conflicting trends: Wu et al.¹⁹ report that convolution-based neural models often outperform fingerprint-based baselines, whereas Mayr et al.²⁰ observe the opposite. These discrepancies arise from differences in dataset construction and evaluation setups, particularly the failure to mimic the distributional shift that occurs in real drug-discovery pipelines. When random data splits are used, GNNs may overfit by memorizing molecular scaffolds rather than learning transferable representations. As a result, models can achieve high test accuracy while failing to generalize to new chemical space. Meaningful comparison of molecular ML methods therefore requires explicit control of scaffold overlap between training and test sets and evaluation protocols designed to reflect realistic generalization requirements.

Table 1 presents a comparative analysis of classical machine-learning methods and chiral-aware Graph Neural Networks, outlining how differences in feature engineering,

adaptability, computational cost, and dataset dependence influence their respective abilities to detect and model molecular chirality.

Table 1 Comparison of Machine Learning and Chiral-Aware Graph Neural Networks

Aspect	Classical Machine Learning	Chiral-Aware Graph Neural Networks
Feature Dependence	Relies heavily on hand-crafted features such as Sterimol parameters and molecular descriptors	Captures molecular topology without predefined features
Performance	Effective in specific domains but faces limitations in extensibility across different chemical reaction spaces	Stronger performance in providing scalable insights into key structural features affecting enantioselectivity
Adaptability	Struggles with novel reaction types due to reliance on predetermined features	Adaptable to new reaction types without requiring manual feature engineering
Data Requirements	Can work with smaller datasets but may require careful feature selection	Requires sufficient and diverse training data
Model Complexity	Often uses simpler algorithms like decision trees or linear regressions, making them easier to interpret	Employs neural networks with multiple layers, captures intricate relationships but can be less interpretable
Interpretability	Easier to interpret with clear physical/chemical parameters	More challenging to interpret compared to traditional ML models
Computational Cost	Lower computation costs and resources	Higher computation costs due to complex architecture
Dataset Size Impact	Small datasets <5k: Classical ML such as SVM, Random Forest, XGBoost show stronger performance for the computational cost	Large datasets >20k: GNN typically outperform classical ML methods

Table 2 Comparative summary of leading chiral-aware graph neural network architectures

Research Paper	Task / Endpoint	Method Used	Baselines Compared	Key Metrics	Main Observation
Tetra-DMPNN ¹⁰	1) Synthetic tasks probing ability to distinguish enantiomers 2) Protein–ligand docking dataset with a chiral pocket	Replaces standard sum aggregation in DMPNN with chiral-sensitive aggregation around tetrahedral centers, using permutations	Vanilla DMPNN with standard sum aggregation	Synthetic tasks: classification accuracy / regression error; docking task: ranking quality	Tetra-DMPNN can distinguish enantiomers and gives modest but consistent improvements over standard message passing on chiral sensitive tasks
ChiRo ¹¹	4 benchmarks: 1. Distinguish conformers of different stereoisomers in latent space. Classify tetrahedral chiral centers as R/S.3. Predict sign of optical rotation (l/d) .4. Rank enantiomers by docking score in a chiral protein pocket.	SE(3)-invariant GNN that embeds sets of coupled torsion angles around bonds with a learned phase-shift, producing rotation about-bond invariance but chirality-sensitive amplitudes. This gives inter-roto invariance & chiral awareness.	2D: DMPNN+chiral tags, Tetra DMPNN (permute/con catenate, with/without tags).3D: SchNet, DimeNet++, SphereNet ¹¹ .	R/S accuracy, l/d accuracy, enantiomer ranking accuracy, contrastive metrics	ChiRo outperforms 2D and 3D baselines on most chiral sensitive benchmarks. It keeps enantiomer clusters separated even under bond rotations due to explicit torsion handling. Where SchNet/DimeNet++ collapse them, and significantly improves l/d and docking ranking accuracy over Tetra-DMPNN and DMPNN+tags
MolKGNN ¹³	QSAR activity prediction on 9 curated high-throughput screening (HTS) datasets (PubChem AIDs) with severe class imbalance (few actives). Target: prioritize hits for drug discovery.	Molecular-kernel GNN that compares local neighborhoods to learnable “molecular kernels”. It is SE(3) invariant, conformation invariant, and chirality-aware, because kernels encode oriented 3D neighborhoods that distinguish enantiomers while being robust to conformer choice.	SchNet, SphereNet, DimeNet++, ChiRo, KerGNN (graph kernel GNN) ¹³	Main: logAUC[0.001, 0.1] (prioritizing top-ranked compounds low FPR) and standard AUC across 9 HTS tasks.	MolKGNN wins or ties on most HTS datasets on the high cutoff logAUC metric and remains competitive on overall AUC. The learned kernels highlight substructures that align with known chiral pharmacophores, making the model both performance leading and interpretable for chiral-driven QSAR.
ChiENN ¹²	A suite of chiral sensitive molecular property prediction tasks, including variants of chiral classification and property prediction where enantiomers share the same 2D graph. Benchmarks target the same failure modes where vanilla GNNs and even torsion-based models struggle.	Order-sensitive message-passing scheme. Neighbors around a node are treated as an ordered ring; messages aggregate over k tuples of consecutive neighbors with a shift-invariant but order-sensitive MLP. No distances/angles/torsions needed; pure orientation of neighbors encodes chirality.	2D: DMPNN, Tetra DMPNN.3D / torsion based: ChiRo.Mode rn GNNs: GPS, SAN, etc., with and without appended ChiENN layers.	Classification / regression metrics (accuracy, ROC-AUC, MAE) across multiple chiral benchmarks;	Adding ChiENN layers consistently outperforms Tetra DMPNN and ChiRo on chiral-sensitive tasks, while remaining domain-agnostic and more scalable (no torsion enumeration, no chiral tags). These results indicate that order-sensitive aggregation around nodes is a promising general framework

Research Paper	Task / Endpoint	Method Used	Baselines Compared	Key Metrics	Main Observation
EnzyKR ¹⁴	Outcomes of enzymatic kinetic resolution (KR): e.g., enantiomeric excess (ee) and conversion for hydrolase-catalyzed resolutions of racemic alcohols/amines – i.e., reaction-level enantioselectivity endpoints.	Chirality-aware reaction model combining molecular graphs and reaction environment descriptors; encodes local chiral environments of substrates and enzymes to predict which enantiomer is favored in KR. (Architecture: GNN blocks specialized for enzyme substrate interactions, with explicit handling of stereocenters.)	Classical ML on descriptors (RF, GBM, MLP), 2D GNNs that ignore chirality or only use chiral tags, and generic 3D GNNs (SchNet-like) without tailored enzyme substrate chirality handling.	Regression metrics on reaction outcomes: R^2 , MAE/RMSE for ee and conversion; classification metrics on “high vs low ee” in some analyses.	EnzyKR substantially improves R^2 and lowers MAE for predicting KR outcomes vs both descriptor-based models and generic GNNs, especially for out-of-substrate domain generalization. Chirality-aware modeling of enzyme substrate complexes is crucial to capturing enantioselectivity trends.
ChiPGT ¹⁸	A set of chirality related benchmarks (similar to ChiRo / ChiENN tasks): • R/S classification. • Enantiomer ranking in binding tasks. • Additional chiral classification/regression datasets constructed to stress enantiomer distinction.	A graph transformer pretrained on large molecular corpora, with chiral-sensitive message passing (building on ideas from Tetra-DMPNN and ChiENN). Uses attention over ordered neighbors or chiral features to encode handedness; fine-tuned on chiral tasks after pretraining.	DMPNN+chiral tags, Tetra-DMPNN, ChiRo, ChiENN, and non pretrained graph transformers.	R/S and enantiomer ranking accuracy; MAE / RMSE for chiral regression tasks. Results are reported both “from scratch” and with pretraining.	Pretraining + chiral aware architecture yields state-of-the-art accuracy on several chirality benchmarks, often surpassing specialized models like ChiRo and ChiENN, especially in low-data regimes. This suggests that chiral structure + large-scale pretraining is a powerful combination for stereochemistry aware modeling.

Table 3 R/S Classification benchmark comparison

Model / Paper (Year)	Dataset	Metric	Performance	Notes
ChiRo ¹¹	PubChem3D R/S dataset	Accuracy	98.5%	Best among torsion-based models
DimeNet++ (2020)	Same	Accuracy	66.3%	Fails to separate enantiomers
SchNet (2017)	Same	Accuracy	54.6%	Collapses mirror structures
2D DMPNN + chiral tags	Same	Accuracy	~75%	Limited by tag expressivity
Tetra-DMPNN ¹⁰	Same	Accuracy	~85–90%	Local ordering improves stereochemical sensitivity
ChiENN ¹²	Multiple datasets	Accuracy	>99%	Order-sensitive aggregation excels
ChiPGT ¹⁸	R/S benchmark	Accuracy	≈99% (pre-trained)	Best performance in low-data regimes

Chiral-aware Neural Network architectures

Conventional GNNs typically use symmetric aggregation operators such as sum, mean, max and are insensitive to chirality, treating enantiomers as indistinguishable due to their identical connectivity. To overcome this gap, a new class of chiral-aware GNN architectures has been developed, integrating 3D positional information, orientation-dependent message passing, and stereochemical constraints into molecular representations.

Table 2 provides a comparative overview of leading chiral-aware GNN approaches and the mechanisms through which they encode stereochemistry for more accurate molecular prediction and design.

Benchmark Results

To evaluate these architectural innovations, we created six task benchmark families focusing on stereochemical understanding. These include assignments of local configuration, global chirality, asymmetric binding, HTS classification, general stereochemical property prediction, enzymatic enantioselectivity, and geometry-complete 3D reasoning. Below, we present a consolidated performance comparison across these benchmarks from the various benchmarks published in the reviewed scientific literature. The datasets used in this review, including those from ChIRo⁸, Tetra-DMPNN⁷, MolKGNN¹⁰, and ChiENN⁹, have been widely used in stereochemistry-aware ML literature. Prior work employing these datasets includes enantiomer ranking (Pattanaik et al., 2020), HTS QSAR benchmarking (Wang et al., 2023), and conformational prediction studies (Adams et al., 2021). The present analysis synthesizes results across these prior benchmarks for comparison.

Task 1: R/S Classification (Absolute Configuration Prediction)

Models must correctly assign the R or S configuration to tetrahedral stereocenters—a fundamental test of chirality expressivity.

As summarized in Table 3, ChiENN, ChIRo, and ChiPGT achieve top-tier accuracy.

Task 2: Optical Rotation (l/d Classification)

Predicting whether a molecule is levorotatory or dextrorotatory requires capturing global, not just local, chirality.

As summarized in Table 4, only ChIRo consistently succeeds. This task highlights the importance of global torsion embedding.

Table 4 Optical Rotation benchmark comparison

Model	Dataset	Metric	Performance
ChIRo (2021)	Optical rotation dataset	Accuracy	79.3%
SphereNet	Same	Accuracy	65.5%
DimeNet++	Same	Accuracy	63–65%
SchNet	Same	Accuracy	55–60%
2D baselines	Same	Accuracy	~50–55%

Task 3: Chiral Docking / Enantiomer Ranking

Accurately ranking enantiomer binding affinities in chiral protein pockets tests a model's ability to encode asymmetric biological interactions.

Table 5 Chiral docking benchmark comparison

Model	Dataset	Metric	Result
ChIRo (2021)	Chiral docking set	Accuracy	> 90%
SchNet / DimeNet++	Same	Accuracy	~70–75%
Tetra-DMPNN	Same	Accuracy	~80–85%
ChiPGT	Same	Accuracy	92–95%
ChiENN	Custom docking tasks	Accuracy	High 90s

As summarized in Table 5, ChiENN, ChIRo, and ChiPGT dominate. Most SE(3)-equivariant models collapse enantiomers and fail.

Task 4: High-Throughput Screening (HTS) QSAR

HTS datasets are noisy, imbalanced, and highly diverse challenging any model's generalizability.

As summarized in Table 6, MolKGNN excels due to kernel-based chirality-aware 3D neighborhoods. Models without explicit chirality representations underperform.

Task 5: General Chiral Property Prediction

Synthetic tasks where molecules differ only in stereochemistry.

As summarized in Table 7, ChiENN provides the most robust cross-task performance.

Table 6 HTS benchmark comparison

Model	Avg logAUC [0.001, 0.1]	Notes
MolKGNN (2023)	0.325	Best early enrichment
SphereNet	0.315	Strong 3D model
DimeNet++	0.312	Good geometry, weak chirality
SchNet	0.282	Collapses enantiomers
ChiRo	0.247	Torsion model struggles with assay noise
KerGNN	0.179	Poor early enrichment

Table 7 General Chiral property benchmark comparison

Model	Metric	Performance	Notes
ChiENN (2023)	Accuracy / MAE	Best on most tasks	Order-sensitive aggregation
ChiRo	Accuracy	80–95%	Good but noise-sensitive
Tetra-DMPNN	Accuracy	70–85%	Limited to tetrahedral centers
SchNet / DimeNet++	Accuracy	50–70%	Often fail to separate enantiomers
ChiPGT	Accuracy	State-of-the-art with pretraining	Excels in low-data tasks

Task 6: Enzymatic Enantioselectivity (EnzyKR)

Predicting enantiomeric excess (ee) and conversion in enzymatic kinetic resolution.

Table 8 Enzymatic ee benchmark comparison

Model	Task	Metric	Performance
EnzyKR (2023)	ee prediction	R ²	0.75–0.80
EnzyKR (2023)	Conversion	RMSE	0.1–0.15
Descriptor models	Ee	R ²	0.50–0.60
Generic GNNs	Ee	R ²	0.55–0.65

As summarized in Table 8, EnzyKR gains from explicit

enzyme–substrate chirality modeling and excels in predicting enantiomeric excess (ee).

Overall Summary

Where available, cross-validation protocols from the original studies were retained, including five-fold CV for ChiRo and MolKGNN, and scaffold splits for ChiENN and Tetra-DMPNN. Metrics reported include AUC, R², RMSE, accuracy, logAUC, and ranking accuracy. This ensures that performance comparisons reflect standardized evaluation. These benchmarks collectively demonstrate that standard GNNs often struggle to model chirality, but modern chiral-aware architectures—ChiRo, Tetra-DMPNN, MolKGNN, ChiENN, ChiPGT, and EnzyKR, offer complementary strengths across stereochemical prediction domains. Among these, ChiENN remains the most generally applicable, while ChiRo excels in torsion-sensitive tasks, MolKGNN in HTS QSAR prediction and EnzyKR in predicting enantiomeric excess.

Discussion**Critical Evaluation of Classical Models and Chiral-Aware Graph Neural Networks**

A key objective of this review is not only to summarize existing approaches but also to critically compare how different machine-learning architectures perform on stereochemistry-dependent prediction tasks. Although the benchmark results show clear performance differences across models, a deeper evaluation reveals important distinctions in expressive capacity, architectural assumptions, data requirements, and generalizability. Together, these factors explain why chiral-aware graph neural networks consistently perform better to classical methods and non-chiral GNNs across nearly all chiral benchmarks.

Classical machine-learning models rely heavily on hand-crafted descriptors such as Sterimol parameters, quantum features, and topological indices. While effective for narrow reaction families, these approaches lack extensibility because new descriptors must be engineered each time a catalyst class changes. More importantly, engineered features typically fail to encode three-dimensional handedness, causing classical models to treat enantiomers as equivalent. This explains their consistently low performance on R/S classification, optical rotation prediction, and enantiomer ranking tasks.

Non-chiral GNNs such as SchNet and DimeNet++ improve molecular representation by learning directly from structure, but their aggregation functions (sum, mean, max) are permutation-invariant. As a result, mirror-image structures produce indistinguishable message-passing updates, causing enantiomers to collapse into the same embedding. This the-

oretical limitation explains why even strong 3D-equivariant models perform near-randomly on chirality-dependent tasks.

In contrast, chiral-aware models succeed because their architectures explicitly incorporate stereochemical information. ChIRo uses torsion-angle sets to encode global chirality and conformational flexibility, making it particularly effective for optical rotation and whole-molecule chirality tasks. ChiENN employs order-sensitive message passing, enabling it to distinguish the spatial arrangement of substituents around local chiral centers without requiring 3D conformers. TetraDMPNN incorporates tetrahedral ordering rules that preserve local stereochemistry, although its applicability is limited to traditional tetrahedral centers. MolKGNN leverages kernel-based 3D neighborhoods, making it robust for noisy high-throughput screening (HTS) tasks even without explicit chirality labels. Finally, pretrained geometric transformers such as ChiPGT achieve strong performance across benchmarks due to their ability to learn both local and global stereochemical cues from large-scale molecular datasets.

Cross-task synthesis highlights several trends. Local chirality tasks (R/S configuration) favor direction-aware models like ChiENN, while global tasks (optical rotation, enantiomeric ranking) require torsion-sensitive architectures like ChIRo or ChiPGT. HTS QSAR tasks benefit from noise-tolerant 3D representations such as MolKGNN. No single architecture performs best across all endpoints, confirming that stereochemistry represents a multidimensional challenge involving both local and global structure.

Despite their advantages, chiral-aware GNNs face remaining challenges, including limited availability of high-quality experimental stereochemical data, computational expense for conformer generation, and the need for unified models that integrate both local and global chirality. Interpretability also remains limited, as even the most accurate models provide minimal insight into which stereochemical features drive predictions. Addressing these limitations will be essential for the next generation of stereochemistry-aware machine learning systems.

Interpretability Challenges

In addition to achieving high predictive accuracy across stereochemistry-sensitive benchmarks, an essential consideration for any machine learning model is interpretability. Interpretability is the ability to relate learned representations to chemically meaningful features such as sterics, torsional strain, electronic distribution, and three-dimensional chirality.

Chiral-aware GNNs differ substantially in their interpretability profiles, depending on whether they rely on torsion embeddings, directional message passing, kernel-based neighborhoods, or full SE(3)-equivariant geometric tensors.

Traditional descriptor-based models were inherently interpretable because their hand-crafted features (e.g., Sterimol parameters, NBO charges) directly mapped onto well-established physical descriptors. To clarify how different feature engineering approaches contribute to enantioselectivity prediction, Table 9 summarizes the main feature categories and their stereochemical impact.

Table 9 Impact of Feature Types on Enantioselectivity

Feature Type	Examples	Impact on Enantioselectivity Prediction
Steric descriptors	Sterimol parameters	Captures substituent bulk, shielding effects
Electronic descriptors	NBO charges, HOMO-LUMO	Models electronic activation, orbital interactions
Torsional features	Dihedral angles, conformer sets	Encodes conformational flexibility, global chirality
3D geometry	Bond distances, angles	Represents spatial arrangement around stereocenters

However, this interpretability came at the cost of poor extensibility and limited generalization. Modern GNN-based architectures reverse this trade-off: they achieve far broader generalization by learning hierarchical representations, but their interpretability must be assessed through their internal mechanisms rather than predefined descriptors.

Interpretability in torsion-based Models (ChIRo)

ChIRo¹¹ offers a natural interpretability pathway because its torsion-set embeddings retain explicit chemical meaning. By grouping torsion angles around common bonds, ChIRo captures conformational flexibility and the energetic consequences of rotation. Analysis of ChIRo's learned embeddings shows that the model separates enantiomers in latent space while maintaining rotational invariance, highlighting torsional patterns that contribute to stereochemical differentiation. This makes ChIRo particularly interpretable for tasks tied to conformational energetics, such as R/S classification or optical rotation prediction.

Interpretability in Chiral Message-Passing Models (TetraDMPNN, ChiENN)

Models that modify the message-passing mechanism provide another layer of interpretability.

- TetraDMPNN¹⁰ uses chirality-guided neighbor orderings, allowing researchers to trace how specific stereo-

center orientations influence aggregated messages. By examining the directional aggregation around a stereocenter, chemists can infer which substituent arrangements drive predicted chiral preferences, for example, why one enantiomer binds more strongly in a protein pocket.

- ChiENN¹² advances interpretability further by basing its aggregation on neighbor directions rather than distances or torsions. This makes the learned features directly connected to local 3D spatial arrangements, offering a geometric rationale for predictions. Importantly, ChiENN's modular nature—acting as a pluggable “chiral-awareness layer”—enables visualization of how chiral asymmetry accumulates across layers in any GNN.

Interpretability in Kernel-Based and Equivariant Frameworks (MolKGNN)

Kernel-based chiral models and geometry-complete equivariant networks gain interpretability through their representation of geometric substructures:

- MolKGNN¹³ learns “molecular kernels” that highlight recurring local patterns associated with chiral biological activity in HTS assays. These kernels can be visualized as chemical fragments whose configurations correlate with activity, yielding insights similar to pharmacophore modeling.

Interpretability Through Pretraining (ChiPGT)

ChiPGT¹⁸ benefits from large-scale pretraining, which allows the model to develop chemically intuitive latent spaces before fine-tuning. These pretrained embeddings often cluster molecules according to stereochemical similarity—behavior consistent with chemists' intuitive organization of chiral scaffolds. This makes ChiPGT's predictions more interpretable via embedding-space visualization (e.g., PCA, t-SNE), particularly for low-data scenarios.

Unresolved Technical Challenges

In addition to interpretability considerations, several unresolved challenges remain across all chiral-aware GNN architectures:

Dependence on High-Quality Conformers

Torsion-based models like ChiRo require accurate 3D conformers, which can be computationally expensive and sensitive to conformer-generation noise. Poor conformers propagate directly into poorer predictions.

Local vs. Global Chirality Integration

Current architecture often specializes in one aspect: ChiENN in local stereocenters; ChiRo in global conformational chirality; MolKGNN in shape-based neighborhoods. Very few models unify local and global chirality, limiting transferability to complex molecular systems.

Data Limitations and Noise Sensitivity

Many chiral datasets are synthetic, computational, or small. Experimental chiral datasets especially with enantioselectivity labels are rare and often noisy. Models such as ChiRo struggle on noisy HTS datasets because chirality signals can be overshadowed by experimental variability.

Scalability and Generalizability

Chiral-aware models have not yet been validated on large, diverse experimental datasets. Their generalization beyond narrowly defined benchmark tasks (e.g., R/S classification or docking) remains uncertain.

Lack of Standardized Stereochemical Benchmarks

The field lacks unified, community-accepted datasets for evaluating chirality-sensitive models. This makes it difficult to compare across studies or ensure reproducibility.

Ethical Challenges

The growing use of AI and generative molecular models in drug discovery and design raises important ethical and responsible-use questions that must be acknowledged in any modern discussion of AI in chemistry.

1. Dual-use Risks in Generative Molecular Design

As generative AI and chiral-aware scoring models advance, the same techniques used to design beneficial pharmaceuticals could theoretically be misused to generate harmful or unethical molecules. AI systems capable of suggesting highly active stereospecific compounds may produce:

- Toxic enantiomers
- Environmentally persistent chiral pollutants
- Molecules with dual-use potential in chemical misuse

Even if unintentional, the ability to rapidly generate novel chiral structures increases the importance of controlled access, responsible publication, and ethical oversight.

2. Bias and Inequities in Chemical Data

Datasets used to train chiral-aware GNNs are often:

-
- Biased toward Western pharmaceutical scaffolds
 - Focused on molecules with commercial relevance
 - Missing stereochemical data for natural products and underserved chemical domains

This may lead to models that disproportionately perform well on certain chemical spaces while ignoring others, reinforcing inequities in drug discovery pipelines. Ensuring diverse chemical representation is necessary for ethical AI deployment.

3. Reproducibility and Scientific Integrity

As chiral-aware GNNs become more powerful, students and researchers may rely on AI predictions without:

- understanding underlying stereochemical mechanisms
- verifying 3D structure assumptions
- checking for dataset leakage or overfitting

Ethical scientific practice requires transparency about model limitations, careful validation, and avoiding overclaiming results.

4. Environmental Impact of Large-Scale Model Training

Pretraining large stereochemistry-sensitive models like ChiPGT or equivariant transformers requires substantial computational resources. High computational energy use contributes to environmental costs. Ethical AI development must weigh:

- model performance
- computational efficiency
- long-term sustainability

5. Responsible Reporting and Open Science

Publishing datasets or model weights without safeguards may unintentionally enable misuse. Conversely, restricting access too heavily limits academic progress. Ethical dissemination requires balanced openness, transparency, and well-defined responsible-use policies.

Conclusion and Future Direction

This review compared various chiral-aware graph neural network architectures to evaluate their effectiveness in predicting stereochemistry-dependent outcomes. By addressing the central research question—*how the choice of model influences*

accuracy, extensibility, and interpretability in enantioselectivity prediction; the synthesis of fourteen peer-reviewed studies reveals a clear trend. Descriptor-based models that use parameters such as Sterimol values and quantum descriptors can achieve strong performance in narrowly defined reaction classes where mechanistic pathways are well understood. However, chiral-aware graph neural networks consistently outperform traditional methods on stereochemistry-sensitive tasks including R/S configuration classification, optical rotation prediction, enantiomer ranking, and chiral HTS QSAR assays. These gains arise from torsion-aware embeddings, order-sensitive message passing, and geometry-complete representations that allow GNNs to learn both local stereocenters and global chiral environments without manual feature engineering.

At the same time, this assessment highlights important limitations and uncertainties that motivate future research. Current architecture depends on diverse, high-quality stereochemical datasets, and no single model dominates across every endpoint. Torsion-based approaches such as ChiRo are sensitive to conformer generation protocols, while order-based architectures like ChiENN primarily model local chirality and may struggle with meso effects or distal interactions. Large pre-trained models such as ChiPGT achieve state-of-the-art performance in low-data settings but require significant computational resources and currently offer limited interpretability compared to classical descriptor-based approaches. Connecting learned representations to physically meaningful stereo-electronic features remains an open challenge that restricts broad adoption.

Future progress in stereochemistry-aware AI will depend on several directions suggested by the literature. Integrating unified local–global chirality modeling could combine the strengths of torsion-driven and order-sensitive approaches, capturing both conformational flexibility and directional substituent effects. Curating open stereochemistry benchmarks—with standardized datasets, metrics, and task definitions—for R/S classification, optical rotation prediction, and enantiomer ranking would improve comparability across studies. Improving interpretability through visualization tools, geometric saliency methods, and attribution analyses could reveal how models capture steric shielding, torsional strain, and noncovalent interactions driving enantioselectivity. Large-scale pretraining has demonstrated potential for low-data domains, and expanding foundation models to include conformer ensembles, catalyst–substrate interactions, and stereochemical supervision may enable broader transfer across reaction classes. Finally, responsible deployment will require attention to dataset bias, transparency in predictive outcomes for safety-relevant compounds, dual-use risks in molecular generation, and sustainable computing practices.

This review consolidates evidence across a rapidly develop-

ing field and highlights that chiral-aware neural architectures represent a promising direction for the design of enantioselective reactions. While classical approaches retain value in settings with limited data or strong mechanistic priors, modern GNNs offer extensible, end-to-end learning of stereochemical effects directly from molecular structure. Advancing stereochemistry-aware AI will require closer integration between machine learning, stereochemical theory, and experimental design to create robust, interpretable, and ethically responsible models that meaningfully accelerate asymmetric catalysis and drug discovery.

References

- 1 N. Pharmaceuticals. Drug discovery and development process. Video]. Available:.
- 2 Chromatographic separation of racemic thalidomide and teratogenic activity of its enantiomers. *Arzneim-Forsch.*
- 3 E. Tokunaga, T. Yamamoto, E. Ito, and N. Shibata. Understanding the thalidomide chirality in biological processes by the self-disproportionation of enantiomers. *Sci. Rep.*, 8(1):17131,.
- 4 C.B. Santiago, A. Milo, and M.S. Sigman. A data-intensive approach to mechanistic elucidation in asymmetric catalysis. *Science*, 347(6227):49–53,.
- 5 M.S. Sigman and E.N. Jacobsen. Quantitative structure–selectivity relationships in asymmetric catalysis. *Chem. Rev.*, 116(18):10087–10124,.
- 6 P. Piras, R. Sheridan, E.C. Sherer, W. Schafer, C.J. Welch, and C. Rousel. Modeling and predicting chiral stationary phase enantioselectivity: An efficient random forest classifier using an optimally balanced training dataset and an aggregation strategy. *J. Sep. Sci.*, 41(6):1365–1375,.
- 7 J. Gilmer, S.S. Schoenholz, P.F. Riley, O. Vinyals, and G.E. Dahl. Neural message passing for quantum chemistry. In *Proceedings of the 34th International Conference on Machine Learning (ICML)*, PMLR, page 1263–1272.
- 8 K. Yang, K. Swanson, W. Jin, C.W. Coley, R. Barzilay, and T. Jaakkola. Analyzing learned molecular representations for property prediction. *J. Mach. Learn. Res.*, 21(202):1–40,.
- 9 J. Gasteiger, J. Gros, M. Gasteiger, and S. Günnemann. Directional message passing for molecular graphs. In *Advances in Neural Information Processing Systems (NeurIPS)*, page 15475–15487.
- 10 L. Pattanaik, O.-E. Ganea, I. Coley, K.F. Jensen, W.H. Green, and C.W. Coley. Message passing networks for molecules with tetrahedral chirality. *arXiv*.
- 11 K. Adams, L. Pattanaik, and C.W. Coley. Learning 3d representations of molecular chirality with invariance to bond rotations. Online]. Available:.
- 12 P. Gaiński, M. Koziarski, J. Tabor, and M. Śmieja. Chienn: Embracing molecular chirality with graph neural networks. Online]. Available:.
- 13 Y. Liu. Interpretable chirality-aware graph neural network for quantitative structure activity relationship modeling in drug discovery. *Proc. AAAI Conf. Artif. Intell.*, 37(12):14356–14364,.
- 14 X. Ran, Y. Jiang, Q. Shao, and Z.J. Yang. Enzykr: a chirality-aware deep learning model for predicting the outcomes of the hydrolase-catalyzed kinetic resolution. *Chem Sci*, 14(43):12073–12082,.
- 15 K.T. Schütt, P.-J. Kindermans, H.E. Sauceda, S. Chmiela, A. Tkatchenko, and K.-R. Müller. Schnet: A continuous-filter convolutional neural network for modeling quantum interactions. In *Advances in Neural Information Processing Systems (NeurIPS)*, page 992–1002.
- 16 J. Klicpera, J. Groß, and S. Günnemann. Fast directional message passing for molecular graphs. In *Advances in Neural Information Processing Systems (NeurIPS)*, page 1633–1646.
- 17 J. Liu, T. Lin, and H. Li. Spherical message passing for 3d molecular graphs. In *International Conference on Learning Representations (ICLR)*.
- 18 Y. Du, J. Chen, and P. Schwaller. Pretrained graph transformers for chirality-related tasks. In *AI for Accelerated Materials Design - Vienna 2024*. Online]. Available:.
- 19 Z. Wu. Moleculenet: a benchmark for molecular machine learning. *Chem. Sci.*, 9(2):513–530,.
- 20 A. Mayr. Large-scale comparison of machine learning methods for drug target prediction on chembl. *Chem. Sci.*, 9(24):5441–5451,.