

Title:

Deep Learning for Blood–Brain Barrier Permeability Prediction

Authors:

Zihan Yang¹, Haipeng Gong^{2*} (Corresponding Author)

Affiliations:

¹Department of Biosciences, Durham University, Durham DH1 3LE, United Kingdom

²School of Life Sciences, Tsinghua University, Beijing, China

*Corresponding author: hgong@tsinghua.edu.cn

Author contact: zihan.yang@durham.ac.uk

Keywords:

Neuropharmacology, Blood–brain barrier permeability, AI in drug discovery, Deep learning, Graph neural networks, Generative modeling

Abstract:

Predicting whether a molecule can cross the blood–brain barrier (*BBB*) is a key step in early-stage neuropharmaceutical development, directly influencing both research efficiency and success rates in drug discovery. Traditional empirical methods based on physicochemical properties are prone to systematic misjudgements due to their reliance on static rules. Early machine learning models, although data-driven, often suffer from limited capacity, poor generalization, and insufficient interpretability. In recent years, artificial intelligence (*AI*)

methods have become essential tools for predicting BBB permeability and guiding related drug design, owing to their ability to model molecular structures and capture complex biological mechanisms. This article systematically reviews the evolution of this field—from deep neural networks to graph-based structural modeling—highlighting the advantages of multi-task and multimodal learning strategies in identifying mechanism-relevant variables. We further explore the emerging potential of generative models and causal inference methods for integrating permeability prediction with mechanism-aware drug design. BBB modeling is in the transition from static classification toward mechanistic perception and structure–function modeling. This paradigm shift provides a methodological foundation and future roadmap for the integration of AI into neuropharmacological development.

Main Text:

Overview of Blood-Brain Barrier Permeability

The development of central nervous system (CNS) drugs has long been characterized by high costs and exceptionally low clinical success rates [1]. Between 2009 and 2018, the clinical failure rate reached 84.6% [2]. One of the principal barriers lies in the blood–brain barrier (BBB)—a highly selective and tightly regulated physiological interface that prevents over 98% of small molecules and nearly all biologics from entering the brain, resulting in frequent clinical-stage attrition of otherwise promising candidates [3–5]. Beyond its physical restrictiveness, the BBB involves a network of complex and incompletely understood transport mechanisms, including passive diffusion, carrier-mediated uptake and energy-dependent efflux, further complicating CNS delivery [6] (see Figure 1 for a schematic summarization [7]). Accordingly, accurate early-stage prediction of BBB permeability is crucial for eliminating non-viable compounds during screening, effectively reducing synthesis and evaluation costs [8]. More importantly, if the modeling of this process

incorporates mechanism-specific representations and path-level semantics, it may provide actionable structural features to support target exposure assessment, rational optimization, and molecular design—bridging the gap between compound structure and clinical accessibility [9,10].

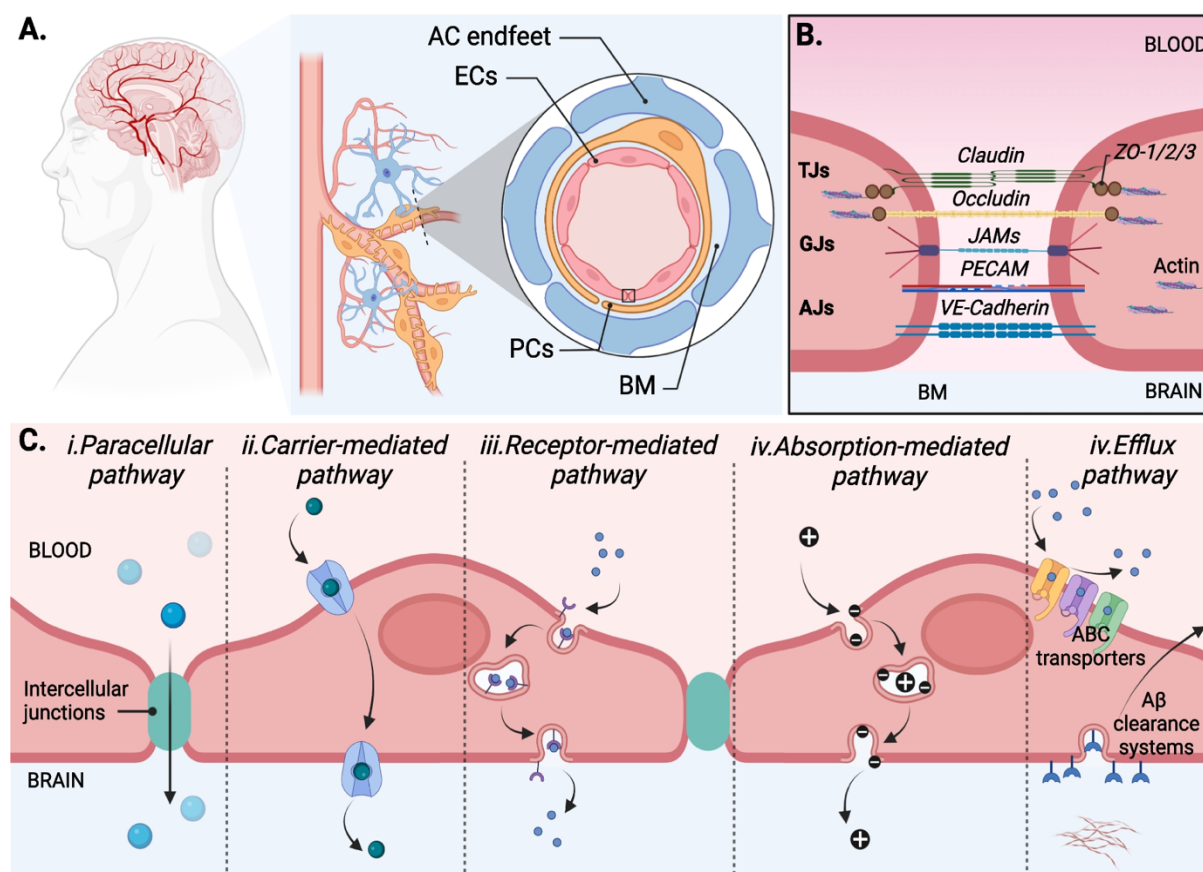


Figure 1. Overview of the BBB structure and transport mechanisms. Adapted from reference [7] under Creative Commons Attribution (CC BY) license.

(A) Neurovascular composition of the BBB, including endothelial cells (ECs), pericytes (PCs), astrocyte endfeet (AC), and the basement membrane (BM).

(B) Junctional complexes between ECs, including tight junctions (TJs), gap junctions (GJs), and adherens junctions (AJs), mediated by claudins, occludin, JAMs, PECAM, and VE-cadherin.

(C) Five key molecular translocation pathways across the BBB:

(i) Paracellular diffusion via intercellular spaces.

- (ii) Carrier-mediated influx via solute carriers.
- (iii) Receptor-mediated endocytosis (e.g., transferrin or insulin);
- (iv) Absorptive transcytosis driven by non-specific uptake.
- (v) Efflux via ABC transporters (e.g., *P*-glycoprotein) responsible for clearing exogenous compounds.

Although deep learning and related AI approaches have been widely introduced to BBB modeling tasks in recent years [11], the majority of existing models continue to formalize the task as a standard classification or regression problem [12]. This framing—driven by sparse mechanistic labels, deployment convenience, and industrial demands for high-throughput compatibility [13]—typically involves learning from molecular descriptors or graph structures to predict a binary indicator of whether a compound can “cross” or not. While this formulation yields reasonably strong predictive performance and is well-suited to screening workflows, it fundamentally reduces a complex, multi-path physiological process to a flat label, obscuring underlying mechanistic heterogeneity. The resulting models often lack interpretability, failing to align semantically with the biological processes they attempt to approximate [14]. Consequently, their outputs provide limited explanatory insight and cannot meaningfully support structure-based design and/or optimization. In response to this structural mismatch, recent studies have begun to explore mechanism-aware modeling, path-variable integration, and even generative approaches for permeability-oriented molecular design [15]—though these efforts remain in early stages. A gradual methodological shift has also emerged: from the global feature-driven deep neural networks (*DNNs*) to the more structurally expressive graph neural networks (*GNNs*) [16,17], and toward generative or mechanism-grounded frameworks. This review adopts the perspective of task definition transformation to systematically examine the evolution of deep learning models for BBB permeability prediction, focusing on their capacity for structural representation, mechanistic

alignment, and actionable feedback, and exploring how future models might bridge the gap between semantic fidelity and practical utility.

Discriminative modeling based on static vector inputs: performance gains amid mechanistic blindness

I. The binary classification framing: historical roots of BBB permeability prediction

The prediction of BBB permeability was historically framed as a binary classification task: determining whether a compound could cross the BBB or not [5]. This framing arose from multiple converging factors shaped by early-stage drug discovery and molecular modeling [3].

First, experimental limitations played a central role. Between the late 20th and early 21st centuries, available methods—such as animal models, *in vitro* permeability assays, and logBB partitioning measurements—primarily yielded coarse, binary outcomes [8], simply classifying compounds as CNS-active or inactive. For instance, classic logBB assays only provided rough estimates of brain-to-blood concentration ratios, without resolving the contributions of passive diffusion versus active transport [13], or detailing molecular interactions with transporters such as *P*-glycoprotein (*P*-gp) and breast cancer resistance protein (*BCRP*) [18]. While modern techniques like single-molecule imaging, transporter-specific knockout models, and microfluidic BBB-on-chip systems now offer richer mechanistic insights [19], such fine-grained data were not accessible when the original task frameworks were established.

Second, the adaptability of computational tools influenced the task definition. Early rule-based systems, physicochemical heuristics, and classical machine learning (*ML*) methods

were inherently aligned with static, fixed-feature classification problems [9]. No mainstream frameworks could process graph-structured or sequential molecular data at the time; models instead relied on precomputed descriptors such as molecular weight, $\log P$ (i.e., logarithm of the octanol–water partition coefficient) and, PSA (i.e., polar surface area) to optimize decision boundaries [10]. These approaches were effective for binary discrimination but lacked the capacity to handle multilevel or dynamic outputs [20].

Last and most importantly, practical industrial needs decisively shaped the task’s formulation. Pharmaceutical companies screening thousands to tens of thousands of compounds in CNS programs needed rapid binary prioritization to eliminate non-viable candidates [1,21], reduce screening costs, and accelerate lead advancement. While mechanistic insights were scientifically valuable, they were deprioritized in favor of throughput and efficiency in early-stage filtering pipelines [22].

The binary framing offered distinct advantages: it lowered data demands, allowed modeling efforts even with small datasets [20], and provided a simple framework for benchmarking algorithmic performance [23]. However, while it initially catalyzed rapid methodological progress, it also embedded structural simplifications that would later hinder mechanistic modeling and design-oriented applications—limitations that are addressed in the subsequent sections [24].

II. Rule-Based and classical ML approaches: early gains and structural constraints

In the development of BBB permeability prediction tasks, rule-based systems and classical ML approaches formed the technological backbone of the early stages [25]. Rule-based methods primarily relied on empirically derived physicochemical thresholds—such as

Lipinski's rule of five, Veber's oral bioavailability criteria, molecular weight cutoffs, $\log P$, and PSA—to link BBB permeability potential with chemical heuristics. These approaches provided simple, interpretable filters widely used in early-stage drug screening to rapidly eliminate unsuitable compounds, especially in settings where only limited experimental permeability data were available [9,26–28].

With the expansion of computational power and data availability, classical *ML* methods such as support vector machine (*SVM*), random forest (*RF*), k-nearest neighbours (*KNN*) and decision tree began to show up in BBB prediction [29,30]. These models typically operated on fixed-input features, including circular molecular fingerprints like the extended-connectivity fingerprints (ECFP) and physicochemical descriptors [20], and used statistical learning to optimize decision boundaries [31]. Compared to rule-based systems, they captured more complex nonlinear relationships and significantly improved predictive accuracy; for example, early SVM-based models achieved notable gains in the area under the curve (*AUC*) of cross-validation and other precision metrics on BBB benchmark datasets [32]. Within the data scale and computational constraints of the time, they provided meaningful performance advances, driving the rapid development of early BBB prediction pipelines [33].

However, these approaches soon revealed inherent structural limitations. First, they were heavily dependent on manually engineered input features, which—while effective for simple chemical space exploration—could not reflect the multilayered nature of molecular structures or only captured local topological patterns [34]. Second, they lacked end-to-end learning capabilities, making it impossible to automatically extract richer patterns (e.g., graph connectivity, three-dimensional conformations, dynamic transport behaviours, etc.) from raw

molecular data [10]. Third, they focused solely on input–output mappings without engaging with the underlying biochemical mechanisms that govern membrane permeation, transporter interactions, and/or energy barriers. In short, they maximized classification performance but offered little mechanistic insight or structural interpretability [35].

Importantly, these limitations were not merely technical constraints but were tightly coupled to the prevailing task definition. As long as the problem was framed as a binary classification task relying on static molecular fingerprints and physicochemical descriptors, classical ML models remained the most practical and efficient solution. Yet, as data complexity increased and research priorities evolved, it became clear that fixed-input, shallow-pattern classification frameworks could no longer meet the growing demand for more complicated and comprehensive modeling objectives [36]—especially those related to mechanistic representation, structural characterization, and multilevel predictions [37].

Against this backdrop, deep learning (DL) methods entered the BBB prediction landscape [38]. Unlike classical ML, deep neural networks (DNNs) offered multilayered feature extraction, hierarchical abstraction, and powerful capacity for learning complex data patterns directly from raw or minimally processed inputs [39]. Their arrival marked not just a performance upgrade but a conceptual shift: from shallow, feature-engineered classification toward deep, end-to-end structural modeling—laying the groundwork for subsequent methodological advances in the field [39].

III. Deep neural networks: feature extraction power and mechanistic limitations

In BBB permeability prediction tasks, DNNs are currently among the best-performing and most widely applied deep learning tools [9]. Compared to traditional ML methods such as

SVM and RF, DNNs offer powerful nonlinear modeling capabilities and multilayer feature extraction, allowing the automatic learning of complex input–output relationships from large-scale molecular datasets [40]. This makes DNNs the standard baseline models in the BBB permeability prediction field [41] and one of the core methods widely adopted in both academic research and industrial applications [13].

The core strength of DNNs lies in two key aspects: the input data and the network architecture. On the input side, DNNs commonly integrate hybrid molecular representations, such as ECFP, *MACCS* key fingerprints, physicochemical descriptors (e.g., $\log P$, molecular weight, hydrogen bond donor/acceptor counts, etc.) and SMILES-based sequence encodings [10]. These representations condense chemical information at multiple levels, capturing not only global molecular properties but also local features such as fragment combinations and neighbourhood environments [42]. For example, ECFP generates multilayer fragment features based on atomic neighbourhoods, while SMILES encodes molecular structures as character sequences, making them suitable for sequence-based modeling [43]. On the architecture side, DNNs primarily rely on multilayer perceptrons (*MLPs*), leveraging techniques such as *ReLU* activation, dropout and batch normalization to efficiently model complex nonlinear relationships in high-dimensional feature spaces [44]. In specific tasks, especially those involving SMILES inputs, some studies have explored using convolutional neural networks (CNNs) to extract local sequence patterns [45] or recurrent neural networks (RNNs) to capture long-range dependencies [46], though these approaches are less common compared to MLPs [44]. Overall, the strength of DNNs comes from the synergy between high-information input features and highly adaptable network architectures, allowing them to process diverse molecular data robustly and effectively.

In practical applications, DNNs have demonstrated outstanding performance across multiple concrete studies. For example, Jing et al. applied ECFP fingerprints with an MLP architecture to ADME-related datasets for BBB permeability prediction, achieving an AUC of ~ 0.90 —significantly outperforming RF- and SVM-based methods [47]. Another study using the *BBBP* dataset employed SMILES encodings combined with a hybrid CNN–DNN multitask model, showing an approximate 6% improvement in accuracy over single-task DNNs, along with enhanced generalizability [44]. Additionally, in industrial high-throughput screening scenarios, there are successful cases where large-scale DNN-based prediction platforms have been used to rapidly screen millions of compounds for potential BBB permeability, significantly reducing the time and cost of experimental validation. These real-world examples highlight not only the academic strength of DNN models but also their practical applicability and scalability in drug development pipelines [48].

However, DNNs also present clear limitations, particularly in the lack of mechanistic interpretation and task extensibility [49]. First, DNNs’ optimization objectives are centered on label fitting, which means that they inherently learn statistical correlations between input features and output labels rather than establishing explicit mechanistic variables or causal pathways [50,51]. Even with the assistance of post hoc explanation tools such as *SHAP* or *LIME* [35,50], DNNs can only provide crude indications on the relative importance of input features, incapable of explaining the underlying questions like “why they matter” or “how they mechanistically affect permeability” [50]. Second, the tight coupling between DNN architectures and their input representations hinders the incorporation of richer dimensions of information, such as molecular topology, three-dimensional conformations, and dynamic transport processes [52]. In particular, due to their reliance on static molecular encodings and label-driven objectives, DNNs are prone to shortcut learning—optimizing for superficial

statistical cues rather than causal molecular features [53]. This deficit also results in poor generalization to out-of-distribution structures, such as novel molecular scaffolds and perturbations in the key functional groups, which are critical for identifying the causal determinants of BBB permeability [54]. Moreover, most current DNN applications remain focused on binary classification tasks (e.g., permeable vs. non-permeable), lacking the capacity to model cross-scale, multitask, or multivariable relationships [55]. These limitations suggest that while DNNs remain strong discriminative tools, they fall short in providing actionable, mechanism-driven insights for molecular optimization and design.

Graph-based structural modeling as a precursor to mechanistic awareness: From semantic mismatch to expressive enhancement

I. Enhancing model interpretability: from attribution analyses to structural embedding

In BBB permeability prediction tasks, model interpretability is not an optional add-on but a core requirement directly tied to drug development [56]. It is not enough to simply predict whether a molecule can cross the BBB; drug developers also need to understand how it happens, which molecular features or mechanisms are responsible, and how structural modifications can improve permeability [57]. However, the DNNs commonly used today are fundamentally black-box models: they provide statistical correlations between input and output but offer no mechanistic explanation [50]. This limitation means that in practical drug development, DNNs often serve only as screening tools, falling short of providing actionable guidance for molecular optimization or mechanistic intervention.

To address this challenge, several efforts have emerged to enhance model interpretability [58]. The first category focuses on shallow interpretability enhancements, applying statistical

attribution analyses over existing inputs. Common tools include SHAP and LIME, which quantify the contribution of input features (such as molecular weight, polarity, and topological indices [59]) to the model's predictions. For example, some studies using SHAP analyses have identified high lipophilicity as a key factor influencing BBB permeability [60], while LIME analyses have shown how the presence of certain molecular fragments can significantly alter model outputs [61]. Additionally, attention mechanisms integrated into SMILES-based deep learning models enable the network to highlight important regions [62,63], such as specific aromatic rings, amine groups, and carboxyl groups. These methods are attractive for their ease of use and intuitive visual outputs, offering a degree of interpretability. However, they share a central limitation: they can only establish correlations between input features and predicted outcomes, without penetrating into the deeper levels of transmembrane mechanisms, conformational dynamics, or local molecular interactions [64]. As a result, the insights they offer have limited value in guiding practical molecular design.

A second category focuses on expanding the diversity and complexity of input information to increase the model's sensitivity to complex mechanisms [65]. Some studies, for example, integrate multiple molecular representations (such as molecular graphs, SMILES encodings, and physicochemical descriptors) into unified models [66], while others employ multitask learning frameworks that jointly optimize permeability predictions alongside pharmacokinetic or *in vitro* experimental data. In one study, incorporating 3D structural and pharmacokinetic data significantly enhanced the model's sensitivity to transport pathways [67]; in another, including membrane transport assay data improved the model's ability to distinguish between different compound classes [68]. However, while these strategies enrich the model's input space, they ultimately remain within the realm of information stitching and do not directly model molecular structure or mechanistic causality [69].

In contrast to these approaches, graph neural networks (GNNs) offer a fundamentally different modeling paradigm [70]. Rather than merely expanding inputs feature or stacking auxiliary labels, GNNs compute directly over molecular graphs, embedding the structural relationships into the computational process [34]. This advantage creates unprecedented opportunities to explore mechanistic variables, local fragments, topological features, and their relationships to functional outcomes [13], which offer a new lens in current interpretability frameworks [71]. To clarify how input representations influence model structure and interpretability, we summarize the comparative characteristics of various modeling strategies in Table 1. This transition—from low-dimensional descriptors to structurally and mechanistically enriched embeddings—marks a progressive redefinition of how permeability prediction is semantically formulated and structurally encoded.

Table 1. Comparative summary of molecular input representation paradigms across structural granularity, modeling capacity, and key limitations.

Input Representation Types		Structural Hierarchy	Dynamic Modeling Capacity	Limitations
Classical Descriptor-based	Physicochemical Descriptors	Molecular Level	Low	Low-dimensional and heuristic-based; Limited generalizability
	Molecular Fingerprints	Sequence Level		Lacks contextual semantics; Fails to capture structural variation effects
Sequence-based	SMILES/Tokenized SMILES	Fragment Level		Token-order sensitive; Lacks contextual and structural coherence
	SMILES Embedding			No graph constraints; Limited spatial and mechanistic expressiveness

Graph-based	2D Molecular Graph	Atom-Level Graph Topology	Weak	No 3D geometry; Lacks stereochemical awareness and cross-fragment generalization
	Graph + Attention		Weak to Moderate	Unstable attribution reliability in attention mechanisms; Supervision-dependent
	Graph + Multimodal			Fusion overhead; Risk of redundancy and semantic mismatch
3D Structure-based	3D Conformational Graph / Atomic Coordinates	3D Spatial Conformation	Moderate to Strong	3D quality-dependent; Unstandardized data; Costly processing
	Spatial Grid / Distance Matrix		Moderate	Sparse and rotation-unstable; Needs standardized preprocessing
Mechanistic Annotation-based	Transporter / Pathway Tags	Mechanism-Annotated Modular Hierarchy	Weak to Moderate	Label-scarce; Annotation burdensome and experiment-dependent
	Causal or Intervention Variables		Moderate	Encoding non-standardized; Data acquisition difficult
	Multi-modal Mechanistic Embeddings		Strong	Representation fragmentation; High data demands

The table contrasts five categories of input representations—classical descriptors, sequence-based, graph-based, 3D structure-based, and mechanistic annotation-based—across structural hierarchy, dynamic modeling capacity, and major limitations.

This framework is synthesized based on current literature in molecular representation learning and mechanistic modeling approaches. [72–74]

II. Structural explanations via GNNs

In response to the growing demand for interpretability, GNNs have emerged in BBB permeability prediction tasks as tools that [75] introduce explicit relationships between nodes, edges and neighbourhoods, albeit unable to fully open the black box of prediction models [76]. This characteristic enables models to at least partially analyze which molecular fragments, local structures, or subgraphs significantly influence the prediction outcomes [77]. Compared to traditional DNNs based on vector inputs, the special attribution capability of GNNs offers a new interpretive perspective [78], potentially layering the foundation for the future mechanism modeling [79].

The key breakthrough of GNNs lies in the design of their input layer: unlike DNNs, GNNs operate directly on molecular graphs, treating atoms as nodes and chemical bonds as edges [80]. Through methods such as graph convolutional networks (GCN) and graph isomorphism networks (GIN), they iteratively aggregate local neighbourhood information [81], integrating hierarchical signals from first-order neighbours (directly bonded atoms) to higher-order neighbours (fragments connected over multiple bonds). For example, under traditional input representations, molecules with identical formulas may still be difficult to distinguish if they differ only in the arrangement of functional groups [9], as fingerprint features often fail to distinguish such nuances. In contrast, GNNs can explicitly identify the spatial positions, chemical environments, and neighbourhood relationships of fragments like aromatic rings, amine groups and carboxyl groups [82], which effectively enhances the resolution of complex molecules, allowing the model to focus on specific local structural signals that are closely related to transmembrane transport or transporter binding. Moreover, beyond recognizing simple local groups, GNNs leverage multi-hop information propagation to comprehensively analyze complex arrangements and interactive effects [83] within molecules. For instance, in molecules with multiple adjacent hydrogen bond donors, GNNs

can integrate the collective influence of these donors to identify potential transmembrane interaction networks, rather than merely calculating the contribution of individual donors. This refined capability allows GNNs to go beyond simply “identifying what the molecule is” to approaching “how the molecule internally operates [84],” offering entirely new perceptual channels for interpretability and subsequent optimization.

In practical applications, the capabilities of GNNs have been validated across multiple studies. The GraphADMET model, using multilayer graph convolutions and multi-hop neighbourhood aggregation, has been shown to detect the significant effects of minor functional group modifications on permeability [85]. Researchers found that simply introducing a methyl or hydroxyl group onto the main chain could lead the model to predict substantial changes in permeability. Notably, such subtle differences are often imperceptible to DNN models during the flattening into global features. MolGNet further incorporates graph attention mechanisms [13], which enables the model to automatically focus on key substructures such as cationic centers, aromatic backbones and flexible single bonds; even when multiple functional regions coexist within a molecule, the model can identify the areas that are the most sensitive to the task output. In a study [86], GNNs were combined with transporter activity predictions, and the results suggested that the model could not only recognize which fragments affect the overall permeability but also pinpoint potential binding sites [87]. For example, in certain molecules, the GNN successfully highlighted amino regions or ring structures likely to interact with *P*-glycoprotein (*P*-gp) transporters [88], providing directly usable, task-relevant clues for molecular optimization. This level of task-level interpretability and attribution not only enhances the comprehensibility of prediction outcomes but also endows GNN models practical utility in drug optimization [86],

particularly in scenarios involving structural adjustments, site selection, and mechanism hypothesis validation.

Despite their notable advances in structural awareness and fragment attribution, the fundamental limitations of GNNs must be clearly acknowledged. First, the optimization objective of GNNs remains constrained to label fitting — whether in classification or regression tasks, the model ultimately learns statistical patterns between input features and output labels, not explicit mechanistic variables [89], causal pathways, or cross-scale processes. The model still cannot explain the detailed biophysical or chemical mechanisms through which local fragments influence permeability [10], nor can it simulate mechanistic changes arising from specific interventions (such as fragment removal, conformational locking, and surface modifications). Second, from a system performance perspective, GNNs place significantly higher demands on data scale, computational resources, and graph preprocessing [90] compared to traditional DNNs. In small-sample or sparsely labelled settings, GNNs are prone to signal dilution and overfitting, while in large-scale high-throughput screening, their complex graph operations and multilayer aggregation mechanisms introduce considerable inference latency and resource consumption. Additionally, the robustness and generalizability of existing GNN architectures across different tasks and/or molecular spaces remain underexplored [91], limiting their widespread deployment in practical drug development workflows.

While GNNs have expanded the structured expression and local interpretability of molecular tasks [10], providing critical preconditions for mechanism modeling, further breakthrough will require more than structural attribution alone [92]. The next stage of research must move toward mechanism-centered frameworks that bridge multiple scales and integrate multi-

module collaborations, truly pushing molecular prediction tasks into a new phase of mechanism-driven, intervention-controllable and optimization-oriented modeling [93].

From prediction to generation: An early leap toward mechanism-aware and integrative modeling

I. Redefining the task: from outcome prediction to mechanistic explanation

In BBB permeability prediction tasks, discriminative modeling frameworks—particularly DNNs—have achieved significant progress, delivering robust performance across key metrics such as accuracy and AUC [94] and becoming important tools for high-throughput screening [95]. Despite these impressive results, such models often serve only as auxiliary references in real-world drug development rather than playing a central decision-making role [96]. The core limitation lies not merely in model performance but in the task definition itself: discriminative tasks focus on optimizing direct input-to-label mappings, without modeling the multilayer variables underlying molecular structure [97], transmembrane mechanisms, or transport pathways. As a result, the outputs provide statistical correlations rather than mechanistic explanations [98], failing to answer essential questions like “Which molecular features drive permeability? [99]” or “How can structural modifications improve outcomes?”

Notably, the demand for mechanistic insight and structural guidance in drug development has existed for a long time [100] but could not be incorporated into task definitions due to multiple constraints: the lack of experimentally annotated mechanistic data, computational limitations [101] in handling complex structural and dynamic information, and industrial priorities that favoured efficiency over interpretability. In recent years, however, these barriers have gradually begun to ease, with the accumulation of experimental data on transporter activities and channel contributions, the enrichment of molecular graph and 3D

conformation databases [102], and the rapid development of computational architectures such as GNNs, generative models, and causal inference frameworks [103]. Although current data volume and integration quality are still insufficient for fully mature mechanistic modeling, the conditions for exploration are increasingly coming into place.

The new task definition is no longer simply about predicting “whether” a molecule can cross the BBB, but about establishing causal chains between molecular structure [104], mechanistic variables, and functional outcomes. For example: Does the molecule cross via passive diffusion or active transport [105]? What functional groups, local conformations, or spatial arrangements play dominant roles [106] in the transmembrane behaviour? If a particular structural region is modified, how will it alter the mechanistic pathway and the ultimate permeability [107]? Such tasks require models to answer not only “what” but also “why” and “how to optimize [108],” providing actionable insights for drug design, screening, and refinement.

Therefore, “mechanistic modeling” is not merely about introducing more sophisticated algorithms—it represents a complete reshaping of the modeling framework [109]. It demands richer input layers (e.g., molecular graphs, 3D dynamic features, etc.[110]), more complex output indicators (e.g., identification of mechanistic variables, inference of causal pathways [111], prediction of intervention outcomes, etc.), and more advanced optimization goals that integrate across modules [112]. This redefinition of task objectives directly drives the development of subsequent technologies—how to use structural awareness, latent variable learning, generative optimization, and causal reasoning [113]to gradually break through the boundaries of traditional discriminative frameworks, opening new mechanism-driven paths for BBB permeability research. This redefinition of modeling objectives also reshapes how

we categorize modeling approaches: no longer solely by algorithmic type, but by their capacity to encode, reason over, and act upon mechanistic variables. Figure 2 offers a comparative taxonomy across this evolving spectrum.

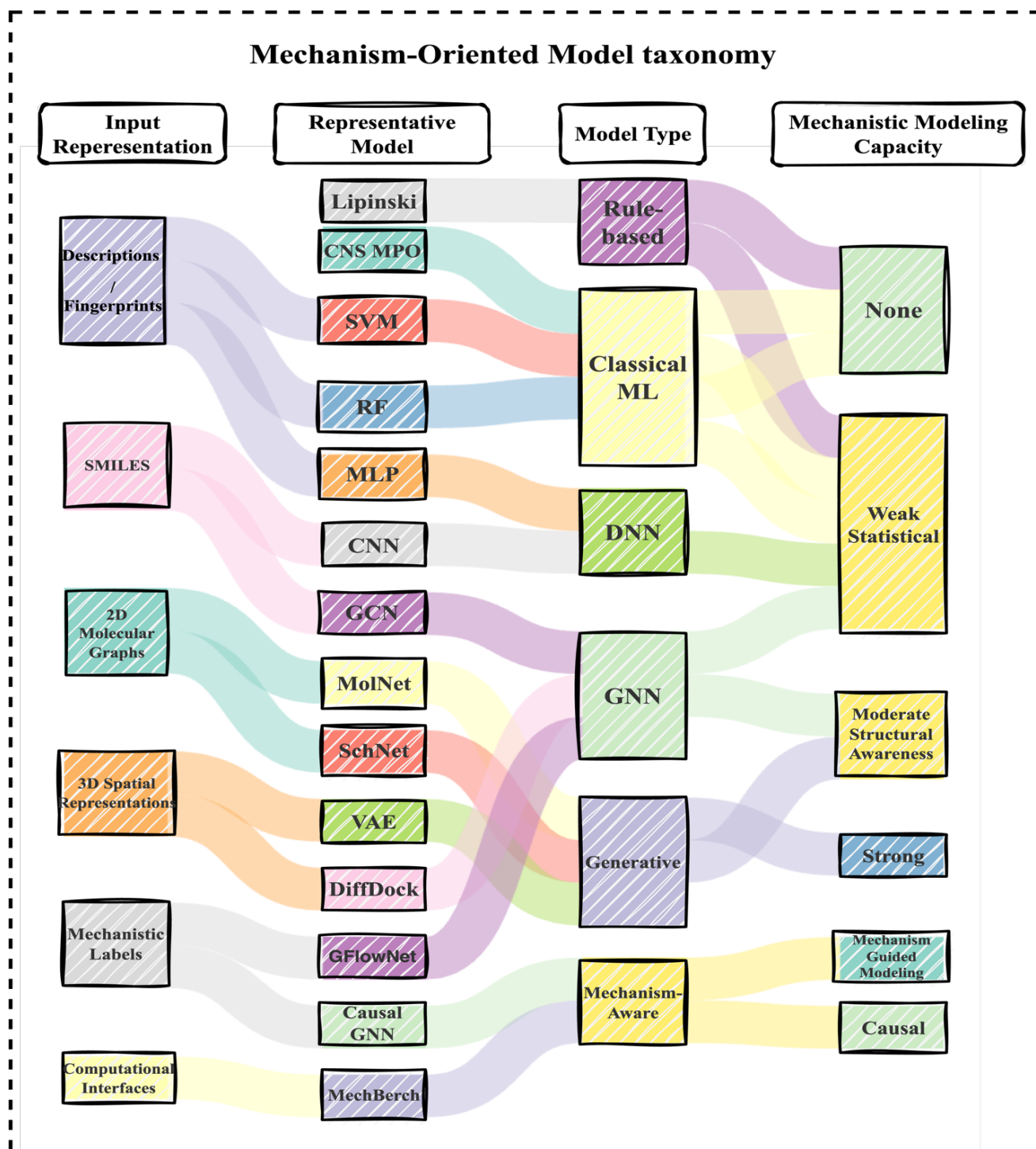


Figure 2. Mechanism-oriented taxonomy of modeling approaches across input modalities, representative architectures, and their capacity for mechanistic reasoning.

Each model type is positioned according to its dominant input representation, typical architectures, and the extent to which it supports mechanistic interpretability—ranging from simple statistical associations to causal modeling. Representative models include classical rule-based and machine learning methods, deep neural networks, graph neural networks, and emerging mechanism-aware frameworks.

II. Current mechanistic modeling approaches: foundational models and representational capabilities

Current mechanistic modeling approaches for BBB permeability prediction cannot be understood as a single unified technique or algorithm [114]. Instead, they represent a multilayered system of interacting components [115], each designed to tackle specific challenges that arise when moving from purely statistical prediction to mechanism-aware reasoning. This layered architecture does not form arbitrarily, but is driven by the inherent complexity of mechanistic modeling itself [116]. To fully grasp it, we must unpack why mechanistic modeling demands distinct functional modules, how these modules are logically connected, and why they must work together as an integrated system.

At its core, mechanistic modeling does not merely aim to establish direct input-output mappings (e.g., predicting whether a molecule crosses the BBB) but requires the explicit introduction and handling of intermediate mechanistic variables—local structural features, dynamic behaviours, transport pathways [117], and/or even latent causal factors—that mediate the relationship between molecular structure and functional outcome. Achieving this goal cannot rely on a single modeling innovation; rather, it necessitates a progressive system where each layer builds up the system's capacity to represent, manipulate and exploit mechanistic information [118,119].

At the foundation, structural modeling components such as GNNs and their three-dimensional extensions (3D GNNs) provide the system with explicit molecular scaffolding [120]. By representing molecules as graphs and applying message-passing mechanisms [121], GNNs capture local substructures and functional group combinations. 3D GNNs further incorporate spatial coordinates and conformational data [122], enabling the model to detect geometric and dynamic features beyond two-dimensional topology [123]. Without this structural foundation, the system lacks the capacity to localize or interpret mechanistic signals [124], making subsequent mechanistic modeling impossible.

Building upon this point, representation learning techniques such as self-supervised and graph contrastive learning [125] become essential for extracting latent mechanistic patterns, especially under conditions of limited data or sparse labelling [126]. These methods employ strategies like masking and reconstruction, positive-negative pair discrimination and cross-modal alignment, allowing the model to autonomously discover mechanism-relevant signals and construct richer [127], mechanism-sensitive representations. The enriched feature space, in turn, equips downstream components with the raw material necessary for mechanism-guided optimization and design.

Advancing further, generative modeling systems such as variational autoencoders (VAEs) and diffusion models [128] transform the system from a passive predictor into an active designer. Rather than merely evaluating known molecules, these models explore latent spaces, generating novel compounds that satisfy mechanistic constraints [129]. This behaviour represents a conceptual shift, reframing the task from screening to design, and enabling the model to contribute directly to hypothesis generation and innovation [130].

At the highest abstraction, causal inference architectures such as causal GNNs and structural causal models [130] embed explicit causal variables, pathways, and intervention reasoning. These models aim not only to learn correlations but to reconstruct causal chains, predict the outcomes of hypothetical interventions [131], and support mechanism-driven scientific inquiry. Although still at an exploratory stage, these methods represent a profound transition from statistical learning to causal reasoning [132].

Importantly, the aforementioned layers are not independent or interchangeable; they form a progressive chain instead. Structural encoding enables mechanistic representation [133]; representation learning enriches generative exploration [134]; generative models supply objects and hypotheses for causal reasoning [135]. The true future breakthroughs in mechanistic modeling will rely not only on advancing individual modules but on integrating them into closed-loop systems, connecting molecular structure, mechanistic understanding, design, and intervention [136]. Such integrated frameworks hold the potential to transform not only BBB permeability prediction but also the broader field of mechanism-driven drug discovery[137], opening new frontiers in both scientific insight and practical application.

III. Toward Practical Implementation: From Mechanistic Modeling to Real-World Application

While the redefinition of the task provides a solid conceptual foundation, realizing its potential in practice demands operational clarity across four key dimensions: data, modeling strategy, evaluation, and application (Figure 3).

- (1) Mechanistic modeling requires data rich in causal and structural context, not just binary outcomes [138]. Key inputs include transporter-specific permeability,

conformational dynamics, annotated transport pathways, as well as proxy information such as physicochemical properties and transition states [139]. These features transform the interpretation of data from outcome-based to mechanism-based. However, most existing datasets lack this level of annotation. To remedy this shortcoming, hybrid strategies—such as simulation-based augmentation, expert-curated transporter knowledge, and causal priors from literature mining—can embed mechanistic signals into model inputs [140], enabling the model to generalize beyond superficial correlations even in low-supervision settings [141].

(2) Mechanism-aware modeling should reflect the hierarchical and causal organization of biological systems [142]. Instead of relying on black-box mappings from molecular descriptors, models should adopt a compositional architecture aligned with mechanistic structure [143]. At the base, multi-task learning leverages auxiliary signals—such as transporter affinity and subcellular localization—to guide the extraction of disentangled, mechanism-relevant features [144]. Causal modeling builds upon this backbone by introducing intermediate variables (e.g., rate-limiting steps, transcytotic routes, etc.), enabling structured reasoning beyond input–output fitting [145]. To address data heterogeneity, multi-fidelity models fuse signals across resolutions—from low-precision estimates to experimental or simulation-derived measurements—ensuring robustness under uncertainty [146]. At the highest layer, physics-informed neural networks (PINNs) incorporate known biophysical laws (e.g., Fick’s diffusion, electrochemical constraints, etc.) directly into training, reinforcing mechanistic grounding [147]. Together, these layers form an integrated scaffold—from supervision to structure and then to constraint—that elevates prediction into biologically coherent reasoning.

(3) Mechanism-aware evaluation is supposed to shift focus from accuracy to causality:

does the model respond in ways consistent with biological mechanisms? [148] At the base, sensitivity tests probe whether the model responds to mechanistic perturbations—such as scaffold changes or activity cliffs—serving as a first-pass filter [144]. Attribution alignment checks whether explanatory signals (e.g., attention, feature importance, etc.) match known drivers [98]. Simulated interventions like transporter knock-outs test whether prediction shifts follow expected causal directions [149]. Causal reasoning tasks go further, using counterfactuals where compound pairs differ only in a key causal factor [150]. Correct differentiation suggests internal causal modeling. At the top, generation tasks evaluate whether models can design compounds that achieve definitive mechanistic goals, such as enhancing transcytosis and selectivity [49]. This hierarchy—sensitivity, attribution, inference, and design—offers a compact yet rigorous framework to assess mechanistic coherence and functional utility [35].

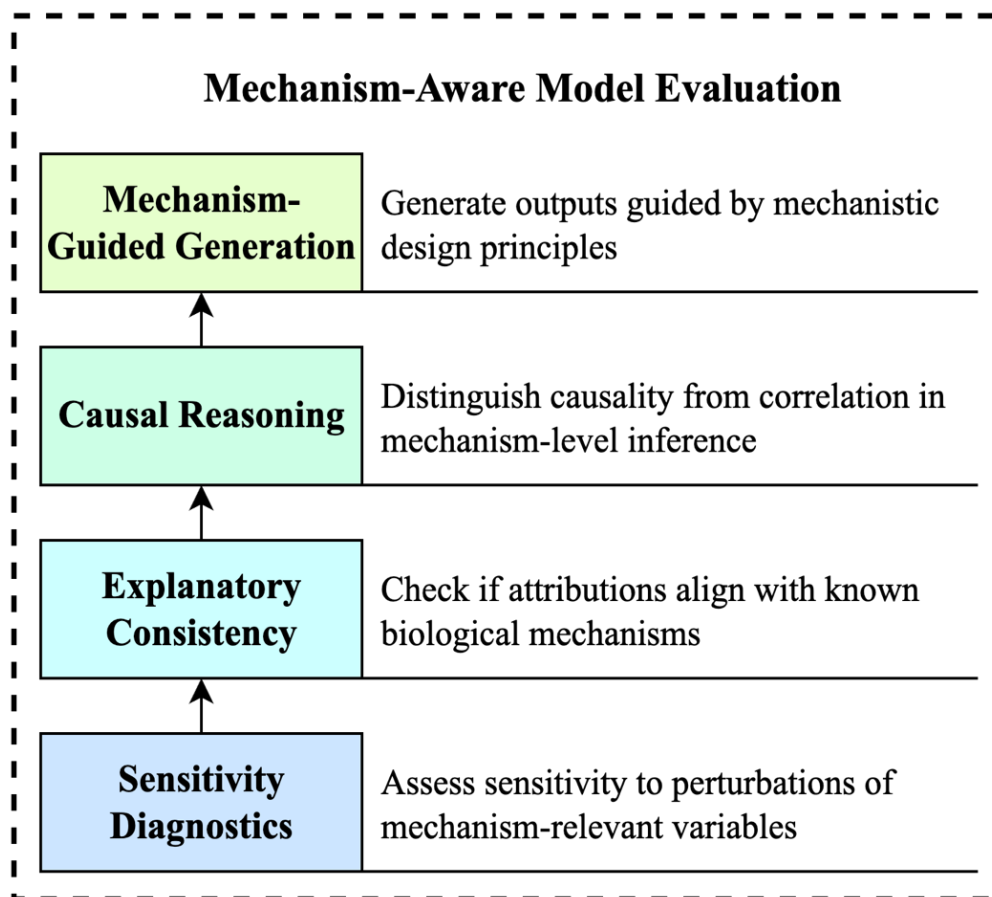


Figure 3. A hierarchical framework for evaluating mechanism-aware models.

The four tiers represent increasing levels of mechanistic alignment:

- (1) Sensitivity diagnostics test response to perturbations;
- (2) Attribution alignment checks correspondence with known drivers;
- (3) Causal reasoning evaluates counterfactual differentiation;
- (4) Mechanism-guided generation assesses goal-directed structural design.

(4) In applications, mechanism-aware systems should serve not just as predictors, but as tools for mechanistic insight [151]. At the base, they enable attribution tracing—linking substructures, functional groups, and spatial motifs to biological effects—supporting targeted optimization (e.g., enhancing BBB transport without compromising efficacy) [152]. Building upward, they simulate counterfactuals and pathway-level interventions: What if a transporter is blocked? What if a scaffold is

modified [149]? These simulations guide rational strategies to exploit or bypass transport routes [151]. At scale, such models support hypothesis generation by mining structure–mechanism associations, potentially revealing overlooked scaffolds or novel transport pathways. In total, these systems shift modeling from outcome prediction to causal reasoning, simulation, and design—bridging molecular observation and mechanism [49].

Taken together, these strategies offer a grounded path for translating mechanism-aware modeling into pharmacological impact. Realizing this promise demands more than algorithmic progress—it requires integrating biological priors, structured mechanistic knowledge, and evaluation frameworks that centered on interpretability and causal utility [140]. Through such integration, models may shift from prediction to explanation, and from data fitting to discovery guidance [35]. This integrative vision is schematized in Figure 4.

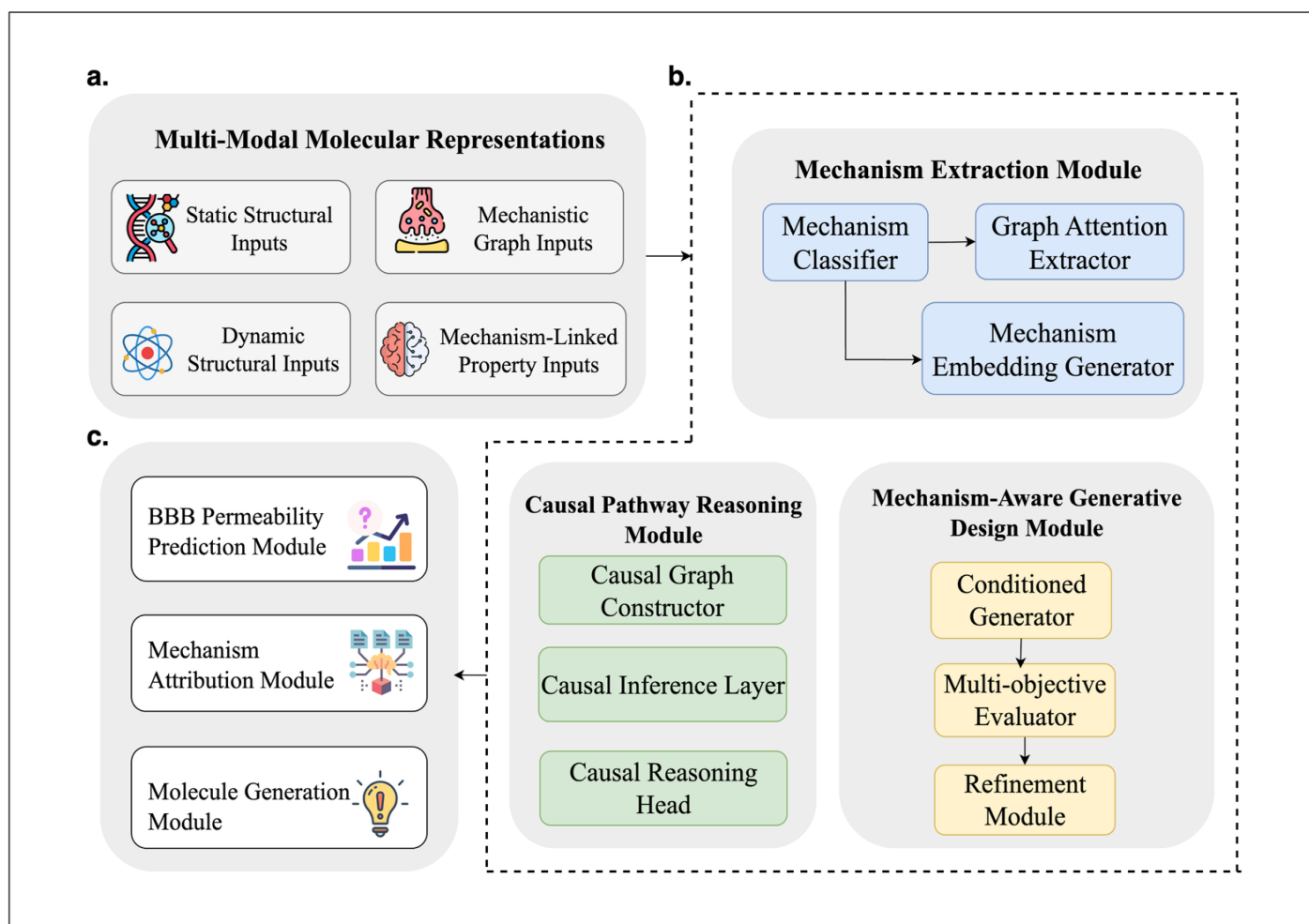


Figure 4. A unified framework for mechanism-aware modeling of BBB permeability and molecular design.

This figure illustrates the architecture of a mechanism-aware modeling framework, composed of three interconnected modules:

- Multi-modal molecular representations integrate static/dynamic structural features, mechanistic graph inputs, and property-linked descriptors to enrich input diversity.
- The mechanism extraction module encodes mechanistic information from inputs via a classifier, graph attention extractor, and embedding generator.
- Three task modules operationalize mechanism-aware modeling: (1) BBB permeability prediction, (2) mechanistic attribution, and (3) mechanism-guided molecular generation.

Two dedicated submodules further enhance mechanistic reasoning: a causal pathway reasoning module that reconstructs causal graphs and inference layers, and a generative design module that optimizes outputs via multi-objective evaluation and refinement.

Current limitations and future directions: challenges and paths toward integrated mechanistic systems

DNNs remain the dominant approach for BBB permeability prediction, balancing strong performance with ease of deployment [153]. Yet their black-box nature limits insight into the structural or causal factors governing trans-barrier behavior [154]. GNNs offer improved molecular representations via structural encoding[9], but remain constrained by discriminative, correlation-based paradigms—falling short of capturing the biological mechanisms underlying permeability [155].

These limitations have made increasingly clear that performance alone is not enough [156]. As predictive models move closer to practical deployment, the emphasis must shift from surface-level correlations to mechanistic interpretability. This need has motivated a growing turn toward mechanism-aware modeling [157]—a paradigm aiming not merely at forecasting outcomes but at reconstructing the causal chains linking molecular structure, transport kinetics, and physiological function. Unlike DNNs and GNNs, which primarily learn statistical associations [157], mechanism-aware approaches explicitly model intermediate variables such as transporter affinity, conformational transitions, and pathway-specific dynamics. These embedded mechanistic representations enable models to explain why a compound is permeable [49], predict how structural changes may modulate its behavior, and generate actionable insights to guide rational molecular design. Grounding prediction in mechanistic understanding will also enhance generalizability across diverse chemical scaffolds, ensure biologically coherent interpretability, and support feedback-driven optimization cycles [140]—capabilities essential for achieving translational impact.

Current mechanistic frameworks remain in the early stages of development and are faced with several persistent challenges, including the scarcity of well-annotated mechanistic datasets, fragmented modeling pipelines, and the absence of unified integration strategies [158]. While foundational tools—such as generative models and causal inference architectures—are beginning to emerge [49], they have yet to form a cohesive system capable of guiding real-world molecular design. Moving forward, the goal is not to replace existing DNN-based infrastructures, but to augment them through a multi-layered, closed-loop architecture that integrates structural encoding, mechanism-specific representation learning, generative molecular design, and causal inference [91]. Such systems will not only enhance predictive accuracy, but also support hypothesis generation, design optimization, and mechanistic interpretability—transforming BBB modeling from statistical forecasting into intervention-guided design [159]. More broadly, this transition reflects a deeper shift in drug discovery itself: from learning patterns to modeling principles and from data-centric predictions to biologically grounded reasoning [140]. This transition, as outlined throughout this review, reflects not only a technical trajectory but a redefinition of how predictive models are framed, evaluated and deployed in neuropharmacological discovery.

Reference:

1. Hay, M. *et al.* (2014) Clinical development success rates for investigational drugs. *Nat. Biotechnol.* 32, 40–51
2. Wong, C.H. *et al.* (2019) Estimation of clinical trial success rates and related parameters. *Biostat. Oxf. Engl.* 20, 273–286
3. (2010) Structure and function of the blood–brain barrier. *Neurobiol. Dis.* 37, 13–25
4. Banks, W.A. (2016) From blood–brain barrier to blood–brain interface: new opportunities for CNS drug delivery. *Nat. Rev. Drug Discov.* 15, 275–292
5. Md, S. *et al.* (2019) Blood-Brain Barrier: From Physiology to Disease and Back. *Physiol. Rev.* 99
6. Daneman, R. and Prat, A. (2015) The Blood–Brain Barrier. *Cold Spring Harb. Perspect. Biol.* 7, a020412
7. Tran, M. *et al.* (2022) Human mini-blood-brain barrier models for biomedical neuroscience research: a review. *Biomater. Res.* 26, 82
8. Pardridge, W.M. (2019) Blood-Brain Barrier and Delivery of Protein and Gene Therapeutics to Brain. *Front. Aging Neurosci.* 11, 373

9. Wu, Z. *et al.* (2018) MoleculeNet: a benchmark for molecular machine learning. *Chem. Sci.* 9, 513–530
10. Yang, K. *et al.* (2019) Analyzing Learned Molecular Representations for Property Prediction. *J. Chem. Inf. Model.* 59, 3370–3388
11. Cornelissen, F.M.G. *et al.* (2023) Explaining Blood–Brain Barrier Permeability of Small Molecules by Integrated Analysis of Different Transport Mechanisms. *J. Med. Chem.* 66, 7253–7267
12. A Bayesian Approach to in Silico Blood-Brain Barrier Penetration Modeling | Journal of Chemical Information and Modeling[Online]. Available: <https://pubs.acs.org/doi/10.1021/ci300124c>. [Accessed: 03-July-2025]
13. Xiong, Z. *et al.* (2020) Pushing the Boundaries of Molecular Representation for Drug Discovery with the Graph Attention Mechanism. *J. Med. Chem.* 63, 8749–8760
14. Jiménez-Luna, J. *et al.* (2020) Drug discovery with explainable artificial intelligence. *Nat. Mach. Intell.* 2, 573–584
15. Zhu, H. *et al.* (2023) A pharmacophore-guided deep learning approach for bioactive molecular generation. *Nat. Commun.* 14, 6234
16. Nguyen, T. *et al.* (2022) Graph Convolutional Networks for Drug Response Prediction. *IEEE/ACM Trans. Comput. Biol. Bioinform.* 19, 146–154
17. Hu, W. *et al.* (2020) Strategies for Pre-training Graph Neural NetworksarXiv
18. Cox, B. *et al.* (2023) The role of the efflux transporter, P-glycoprotein, at the blood-brain barrier in drug discovery. *Biopharm. Drug Dispos.* 44, 113–126
19. Park, T.-E. *et al.* (2019) Hypoxia-enhanced Blood-Brain Barrier Chip recapitulates human barrier function and shuttling of drugs and antibodies. *Nat. Commun.* 10, 2621
20. Muehlbacher, M. *et al.* (2011) Qualitative prediction of blood-brain barrier permeability on a large and refined dataset. *J. Comput. Aided Mol. Des.* 25, 1095–1106
21. Blood-brain barrier-on-a-chip: Microphysiological systems that capture the complexity of the blood-central nervous system interface - PubMed[Online]. Available: <https://pubmed.ncbi.nlm.nih.gov/28195514/>. [Accessed: 06-July-2025]
22. Wager, T.T. *et al.* (2016) Central Nervous System Multiparameter Optimization Desirability: Application in Drug Discovery. *ACS Chem. Neurosci.* 7, 767–775
23. Meng, F. *et al.* (2021) A curated diverse molecular database of blood-brain barrier permeability with chemical descriptors. *Sci. Data* 8, 289
24. Explaining Blood–Brain Barrier Permeability of Small Molecules by Integrated Analysis of Different Transport Mechanisms | Journal of Medicinal Chemistry[Online]. Available: https://pubs.acs.org/doi/10.1021/acs.jmedchem.2c01824?utm_source=chatgpt.com. [Accessed: 03-July-2025]
25. Rankovic, Z. (2015) CNS drug design: balancing physicochemical properties for optimal brain exposure. *J. Med. Chem.* 58, 2584–2608
26. Central Nervous System Multiparameter Optimization Desirability: Application in Drug Discovery | ACS Chemical Neuroscience[Online]. Available: <https://pubs.acs.org/doi/10.1021/acscemneuro.6b00029>. [Accessed: 03-July-2025]
27. Veber, D.F. *et al.* (2002) Molecular Properties That Influence the Oral Bioavailability of Drug Candidates. *J. Med. Chem.* 45, 2615–2623
28. (1997) Experimental and computational approaches to estimate solubility and permeability in drug discovery and development settings. *Adv. Drug Deliv. Rev.* 23, 3–25
29. Zhao, Y.H. *et al.* (2007) Predicting penetration across the blood-brain barrier from simple descriptors and fragmentation schemes. *J. Chem. Inf. Model.* 47, 170–175

30. Vatansever, S. *et al.* (2021) Artificial intelligence and machine learning-aided drug discovery in central nervous system diseases: State-of-the-arts and future directions. *Med. Res. Rev.* 41, 1427–1473
31. Wang, Z. *et al.* (2018) In Silico Prediction of Blood-Brain Barrier Permeability of Compounds by Machine Learning and Resampling Methods. *ChemMedChem* 13, 2189–2201
32. Martins, I.F. *et al.* (2012) A Bayesian approach to in silico blood-brain barrier penetration modeling. *J. Chem. Inf. Model.* 52, 1686–1697
33. Moving beyond Rules: The Development of a Central Nervous System Multiparameter Optimization (CNS MPO) Approach To Enable Alignment of Druglike Properties | ACS Chemical Neuroscience[Online]. Available: <https://pubs.acs.org/doi/10.1021/cn100008c>. [Accessed: 03-July-2025]
34. Gilmer, J. *et al.* (2017) Neural Message Passing for Quantum Chemistry. in *Proceedings of the 34th International Conference on Machine Learning*, pp. 1263–1272
35. Doshi-Velez, F. and Kim, B. (2017) Towards A Rigorous Science of Interpretable Machine LearningarXiv
36. Zeng, X. *et al.* (2022) Deep generative molecular design reshapes drug discovery. *Cell Rep. Med.* 3, 100794
37. Deep Learning in Drug Discovery - Gawehn - 2016 - Molecular Informatics - Wiley Online Library[Online]. Available: <https://onlinelibrary.wiley.com/doi/10.1002/minf.201501008>. [Accessed: 03-July-2025]
38. Gao, K. *et al.* (2020) Generative Network Complex for the Automated Generation of Drug-like Molecules. *J. Chem. Inf. Model.* 60, 5682–5698
39. Duvenaud, D.K. *et al.* (2015) Convolutional Networks on Graphs for Learning Molecular Fingerprints. in *Advances in Neural Information Processing Systems*, 28
40. Sharma, A. *et al.* (2024) ANN multi-layer perceptron for prediction of blood-brain barrier permeable compounds for central nervous system therapeutics. *J. Biomol. Struct. Dyn.* DOI: 10.1080/07391102.2024.2326671
41. Mayr, A. *et al.* (2016) DeepTox: Toxicity Prediction using Deep Learning. *Front. Environ. Sci.* 3
42. Extended-Connectivity Fingerprints | Journal of Chemical Information and Modeling[Online]. Available: <https://pubs.acs.org/doi/10.1021/ci100050t>. [Accessed: 03-July-2025]
43. Automatic Chemical Design Using a Data-Driven Continuous Representation of Molecules | ACS Central Science[Online]. Available: <https://pubs.acs.org/doi/10.1021/acscentsci.7b00572>. [Accessed: 03-July-2025]
44. Altae-Tran, H. *et al.* (2016) Low Data Drug Discovery with One-shot LearningarXiv
45. Kimber, T.B. *et al.* (2018) Synergy Effect between Convolutional Neural Networks and the Multiplicity of SMILES for Improvement of Molecular PredictionarXiv
46. Molecular Transformer: A Model for Uncertainty-Calibrated Chemical Reaction Prediction | ACS Central Science[Online]. Available: <https://pubs.acs.org/doi/10.1021/acscentsci.9b00576>. [Accessed: 03-July-2025]
47. GCN-BBB: Deep Learning Blood-Brain Barrier (BBB) Permeability PharmacAnalytics with Graph Convolutional Neural (GCN) Network | The AAPS Journal[Online]. Available: <https://link.springer.com/article/10.1208/s12248-025-01059-0>. [Accessed: 03-July-2025]
48. Xia, J. *et al.* Understanding the Limitations of Deep Models for Molecular property prediction: Insights and Solutions
49. Sanchez-Lengeling, B. and Aspuru-Guzik, A. (2018) Inverse molecular design using machine learning: Generative models for matter engineering. *Science* 361, 360–365

50. Rudin, C. (2019) Stop explaining black box machine learning models for high stakes decisions and use interpretable models instead. *Nat. Mach. Intell.* 1, 206–215
51. Caldas Ramos, M. *et al.* (2025) A review of large language models and autonomous agents in chemistry. *Chem. Sci.* 16, 2514–2572
52. Méndez-Lucio, O. *et al.* (2021) A geometric deep learning approach to predict binding conformations of bioactive molecules. *Nat. Mach. Intell.* 3, 1033–1039
53. Geirhos, R. *et al.* (2020) Shortcut learning in deep neural networks. *Nat. Mach. Intell.* 2, 665–673
54. Hu*, W. *et al.* (2019) Strategies for Pre-training Graph Neural Networks. presented at the International Conference on Learning Representations
55. Analyzing Learned Molecular Representations for Property Prediction | Journal of Chemical Information and Modeling[Online]. Available: <https://pubs.acs.org/doi/10.1021/acs.jcim.9b00237>. [Accessed: 03-July-2025]
56. (PDF) Drug discovery with explainable artificial intelligence. *ResearchGate* DOI: 10.1038/s42256-020-00236-4
57. A BOILED-Egg To Predict Gastrointestinal Absorption and Brain Penetration of Small Molecules - PubMed[Online]. Available: <https://pubmed.ncbi.nlm.nih.gov/27218427/>. [Accessed: 06-July-2025]
58. Gallegos, M. *et al.* (2024) Explainable chemical artificial intelligence from accurate machine learning of real-space chemical descriptors. *Nat. Commun.* 15, 4345
59. Lundberg, S. and Lee, S.-I. (2017) A Unified Approach to Interpreting Model PredictionsarXiv
60. Practical guide to SHAP analysis: Explaining supervised machine learning model predictions in drug development - Ponce-Bobadilla - 2024 - Clinical and Translational Science - Wiley Online Library[Online]. Available: <https://ascpt.onlinelibrary.wiley.com/doi/full/10.1111/cts.70056>. [Accessed: 03-July-2025]
61. Tsubaki, M. *et al.* (2019) Compound-protein interaction prediction with end-to-end learning of neural networks for graphs and sequences. *Bioinforma. Oxf. Engl.* 35, 309–318
62. (2024) Transformer-based models for chemical SMILES representation: A comprehensive literature review. *Heliyon* 10, e39038
63. Schwaller, P. *et al.* (2019) Molecular Transformer: A Model for Uncertainty-Calibrated Chemical Reaction Prediction. *ACS Cent. Sci.* 5, 1572–1583
64. Deep Learning: A Critical Appraisal[Online]. Available: https://www.researchgate.net/publication/322220859_Deep_Learning_A_Critical_Appraisal. [Accessed: 06-July-2025]
65. (2018) The rise of deep learning in drug discovery. *Drug Discov. Today* 23, 1241–1250
66. Ramsundar, B. *et al.* (2017) Is Multitask Deep Learning Practical for Pharma? *J. Chem. Inf. Model.* 57, 2068–2076
67. Myung, Y. *et al.* (2024) Deep-PK: deep learning for small molecule pharmacokinetic and toxicity prediction. *Nucleic Acids Res.* 52, W469–W475
68. Sedykh, A. *et al.* (2013) Human Intestinal Transporter Database: QSAR Modeling and Virtual Profiling of Drug Uptake, Efflux and Interactions. *Pharm. Res.* 30, 996–1007
69. Li, R. *et al.* (2024) Physical formula enhanced multi-task learning for pharmacokinetics predictionarXiv
70. Wu, Z. *et al.* (2021) A Comprehensive Survey on Graph Neural Networks. *IEEE Trans. Neural Netw. Learn. Syst.* 32, 4–24
71. Ying, Z. *et al.* (2019) GNNExplainer: Generating Explanations for Graph Neural Networks. in *Advances in Neural Information Processing Systems*, 32

72. Jumper, J. *et al.* (2021) Highly accurate protein structure prediction with AlphaFold. *Nature* 596, 583–589
73. Gawehn, E. *et al.* (2016) Deep Learning in Drug Discovery. *Mol. Inform.* 35, 3–14
74. Gilmer, J. *et al.* (2017) Neural Message Passing for Quantum ChemistryarXiv
75. Relational graph convolutional networks for predicting blood–brain barrier penetration of drug molecules | Bioinformatics | Oxford Academic[Online]. Available: <https://academic.oup.com/bioinformatics/article/38/10/2826/6564684>. [Accessed: 04-July-2025]
76. Baldassarre, F. and Azizpour, H. (2019) Explainability Techniques for Graph Convolutional NetworksarXiv
77. Sun R. *et al.* (2022) Does GNN Pretraining Help Molecular Representation? *Adv. Neural Inf. Process. Syst.* 35, 12096–12109
78. HimGNN: a novel hierarchical molecular graph representation learning framework for property prediction | Briefings in Bioinformatics | Oxford Academic[Online]. Available: <https://academic.oup.com/bib/article/24/5/bbad305/7245716>. [Accessed: 04-July-2025]
79. Sanchez-Lengeling, B. *et al.* (2020) Evaluating Attribution for Graph Neural Networks. in *Advances in Neural Information Processing Systems*, 33, pp. 5898–5910
80. Zang, X. *et al.* (2023) Hierarchical Molecular Graph Self-Supervised Learning for property prediction. *Commun. Chem.* 6, 34
81. Graph neural networks for materials science and chemistry - PMC[Online]. Available: <https://pmc.ncbi.nlm.nih.gov/articles/PMC9702700/>. [Accessed: 04-July-2025]
82. Integrating concept of pharmacophore with graph neural networks for chemical property prediction and interpretation | Journal of Cheminformatics | Full Text[Online]. Available: <https://jcheminf.biomedcentral.com/articles/10.1186/s13321-022-00634-3>. [Accessed: 04-July-2025]
83. Wu, Z. *et al.* (2023) Chemistry-intuitive explanation of graph neural networks for molecular property prediction with substructure masking. *Nat. Commun.* 14, 2585
84. Molecule generation for drug design: A graph learning perspective - ScienceDirect[Online]. Available: <https://www.sciencedirect.com/science/article/pii/S2667325824005259>. [Accessed: 04-July-2025]
85. Kipf, T.N. and Welling, M. (2017) Semi-Supervised Classification with Graph Convolutional Networks. presented at the International Conference on Learning Representations
86. Gasteiger, J. *et al.* (2019) Directional Message Passing for Molecular Graphs. presented at the International Conference on Learning Representations
87. Modeling polypharmacy side effects with graph convolutional networks | Bioinformatics | Oxford Academic[Online]. Available: <https://academic.oup.com/bioinformatics/article/34/13/i457/5045770>. [Accessed: 06-July-2025]
88. (PDF) Natural compounds from Ginkgo biloba L. targeting MPO, ERAP-2, DPP-4, PPAR- α , and ET B in cardiovascular diseases: Molecular docking and ADMET predictions[Online]. Available: https://www.researchgate.net/publication/382292542_Natural_compounds_from_Ginkgo_biloba_L_targeting_MPO_ERAP-2_DPP-4_PPAR-alpha_and_ET_B_in_cardiovascular_diseases_Molecular_docking_and_ADMET_predictions. [Accessed: 04-July-2025]
89. Duvenaud, D. *et al.* (2015) Convolutional Networks on Graphs for Learning Molecular FingerprintsarXiv
90. Dwivedi, V.P. *et al.* (2022) Benchmarking Graph Neural NetworksarXiv

91. Graph neural networks: A review of methods and applications - ScienceDirect[Online]. Available: <https://www.sciencedirect.com/science/article/pii/S2666651021000012>. [Accessed: 04-July-2025]
92. El, B. *et al.* (2025) Towards Mechanistic Interpretability of Graph Transformers via Attention Graphs *arXiv*
93. Ahmedt-Aristizabal, D. *et al.* (2021) Graph-Based Deep Learning for Medical Diagnosis and Analysis: Past, Present and Future. *Sensors* 21, 4758
94. Feinberg, E.N. *et al.* (2018) PotentialNet for Molecular Property Prediction. *ACS Cent. Sci.* 4, 1520–1530
95. CNSMolGen: A Bidirectional Recurrent Neural Network-Based Generative Model for De Novo Central Nervous System Drug Design | Journal of Chemical Information and Modeling[Online]. Available: <https://pubs.acs.org/doi/10.1021/acs.jcim.4c00504>. [Accessed: 04-July-2025]
96. Blanco-González, A. *et al.* (2023) The Role of AI in Drug Discovery: Challenges, Opportunities, and Strategies. *Pharmaceuticals* 16, 891
97. Zagidullin, B. *et al.* (2021) Comparative analysis of molecular fingerprints in prediction of drug combination effects. *Brief. Bioinform.* 22, bbab291
98. Ribeiro, M.T. *et al.* (2016) “Why Should I Trust You?”: Explaining the Predictions of Any Classifier *arXiv*
99. (PDF) Data science and molecular biology: prediction and mechanistic explanation[Online]. Available: https://www.researchgate.net/publication/333519272_Data_science_and_molecular_biology_prediction_and_mechanistic_explanation. [Accessed: 04-July-2025]
100. Gill, J. *et al.* (2022) Comparing the applications of machine learning, PBPK, and population pharmacokinetic models in pharmacokinetic drug–drug interaction prediction. *CPT Pharmacomet. Syst. Pharmacol.* 11, 1560–1568
101. Altae-Tran, H. *et al.* (2017) Low Data Drug Discovery with One-Shot Learning. *ACS Cent. Sci.* 3, 283–293
102. Lee, S. *et al.* (2024) Molecule Generation with Fragment Retrieval Augmentation *arXiv*
103. Moret, M. *et al.* (2023) Leveraging molecular structure and bioactivity with chemical language models for de novo drug design. *Nat. Commun.* 14, 114
104. Yuan, T. *et al.* (2025) Decoding Brain Interstitial Transport In Vivo: A Fully Validated Bottom-Up Mechanistic Prediction Framework *bioRxiv*, 2025.03.22.643151
105. Permeability across lipid membranes - ScienceDirect[Online]. Available: <https://www.sciencedirect.com/science/article/pii/S0005273616301316>. [Accessed: 04-July-2025]
106. Pardridge, W.M. (2012) Drug transport across the blood-brain barrier. *J. Cereb. Blood Flow Metab. Off. J. Int. Soc. Cereb. Blood Flow Metab.* 32, 1959–1972
107. Kadry, H. *et al.* (2020) A blood–brain barrier overview on structure, function, impairment, and biomarkers of integrity. *Fluids Barriers CNS* 17, 69
108. Brown, N. *et al.* (2019) GuacaMol: Benchmarking Models for de Novo Molecular Design. *J. Chem. Inf. Model.* 59, 1096–1108
109. Schölkopf, B. *et al.* (2021) Toward Causal Representation Learning. *Proc. IEEE* 109, 612–634
110. Bayesian probabilistic approach for predicting backbone structures in terms of protein blocks | Request PDF. *ResearchGate* at https://www.researchgate.net/publication/353480793_Bayesian_probabilistic_approach_for_predicting_backbone_structures_in_terms_of_protein_blocks

111. (PDF) Graph Structure Learning for Robust Graph Neural Networks[Online]. Available: https://www.researchgate.net/publication/341539285_Graph_Structure_Learning_for_Robust_Graph_Neural_Networks. [Accessed: 06-July-2025]
112. Zhao, J. *et al.* (2024) Exploring a Small Molecule Property Prediction Model with Optimal Comprehensive Performance through Multi-Objective Optimization Algorithms. in *Proceedings of the Genetic and Evolutionary Computation Conference Companion*, New York, NY, USA, pp. 707–710
113. Conditional Molecular Design with Deep Generative Models | Request PDF. *ResearchGate* DOI: 10.1021/acs.jcim.8b00263
114. Bridging the gap between mechanistic biological models and machine learning surrogates | PLOS Computational Biology[Online]. Available: <https://journals.plos.org/ploscompbiol/article?id=10.1371/journal.pcbi.1010988>. [Accessed: 07-July-2025]
115. Jiménez-Luna, J. *et al.* (2020) Drug discovery with explainable artificial intelligencearXiv
116. Harris, C.R. *et al.* (2020) Array programming with NumPy. *Nature* 585, 357–362
117. Zhang, S. *et al.* (2022) Mechanistic Insights into Passive Membrane Permeability of Drug-like Molecules from a Weighted Ensemble of Trajectories. *J. Chem. Inf. Model.* 62, 1891–1904
118. Zeng, X. *et al.* (2022) Accurate prediction of molecular properties and drug targets using a self-supervised image representation learning framework. *Nat. Mach. Intell.* 4, 1004–1016
119. Qiao, L. *et al.* (2024) The evolution of systems biology and systems medicine: From mechanistic models to uncertainty quantificationarXiv
120. Maziarz, K. *et al.* (2024) Learning to Extend Molecular Scaffolds with Structural MotifsarXiv
121. Han, X. *et al.* (2025) Integrated multimodal hierarchical fusion and meta-learning for enhanced molecular property prediction. *Brief. Bioinform.* 26, bbaf251
122. Han, C. *et al.* (2024) Accurate Prediction of NMR Chemical Shifts: Integrating DFT Calculations with Three-Dimensional Graph Neural Networks. *J. Chem. Theory Comput.* 20, 5250–5258
123. Batzner, S. *et al.* (2022) E(3)-equivariant graph neural networks for data-efficient and accurate interatomic potentials. *Nat. Commun.* 13, 2453
124. (2020) Vertex-frequency graph signal processing: A comprehensive review. *Digit. Signal Process.* 107, 102802
125. Li, S. *et al.* (2021) GeomGCL: Geometric Graph Contrastive Learning for Molecular Property PredictionarXiv
126. (2023) Few-shot learning with transformers via graph embeddings for molecular property prediction. *Expert Syst. Appl.* 225, 120005
127. Zheng, S. *et al.* (2022) Cross-modal Graph Contrastive Learning with Cellular Images. at <https://openreview.net/forum?id=V_V53-WG9m6>
128. Kusner, M.J. *et al.* (2017) Grammar Variational Autoencoder. in *Proceedings of the 34th International Conference on Machine Learning*, pp. 1945–1954
129. (2024) Unleashing the power of generative AI in drug discovery. *Drug Discov. Today* 29, 103992
130. Sadybekov, A.V. and Katritch, V. (2023) Computational approaches streamlining drug discovery. *Nature* 616, 673–685
131. (2023) Interventions in Graph Neural Networks Lead to New Neural Causal Models. *Trans. Mach. Learn. Res.* at <<https://openreview.net/forum?id=PzdnLSvyiJ¬eId=tJkgXCXkgU>>

132. Rogers, J.A. *et al.* (2023) An introduction to causal inference for pharmacometricians. *CPT Pharmacomet. Syst. Pharmacol.* 12, 27–40
133. Weller, J.A. and Rohs, R. (2024) Structure-Based Drug Design with a Deep Hierarchical Generative Model. *J. Chem. Inf. Model.* 64, 6450–6463
134. (2022) Generative machine learning for de novo drug discovery: A systematic review. *Comput. Biol. Med.* 145, 105403
135. Molecule Design by Latent Prompt Transformer[Online]. Available: <https://arxiv.org/html/2402.17179v2>. [Accessed: 04-July-2025]
136. Ivanenkov, Y.A. *et al.* (2023) Chemistry42: An AI-Driven Platform for Molecular Design and Optimization. *J. Chem. Inf. Model.* 63, 695–701
137. Jiang, F. *et al.* (2017) Artificial intelligence in healthcare: past, present and future. *Stroke Vasc. Neurol.* 2, 230–243
138. Schölkopf, B. *et al.* (2021) Towards Causal Representation LearningarXiv
139. Ruck, T. *et al.* (2015) Blood-brain barrier modeling: challenges and perspectives. *Neural Regen. Res.* 10, 889–891
140. (2019) Machine learning for integrating data in biology and medicine: Principles, practice, and opportunities. *Inf. Fusion* 50, 71–91
141. Chen, Y.-S. and Shuai, H.-H. (2021) Meta-Transfer Learning for Low-Resource Abstractive Summarization. *arXiv.org*. [Online]. Available: <https://arxiv.org/abs/2102.09397v2>. [Accessed: 21-July-2025]
142. Alon, U. (2007) Network motifs: theory and experimental approaches. *Nat. Rev. Genet.* 8, 450–461
143. Battaglia, P.W. *et al.* (2018) Relational inductive biases, deep learning, and graph networksarXiv
144. Amaro, R.E. and Mulholland, A.J. (2018) Multiscale Methods in Drug Design Bridge Chemical and Biological Complexity in the Search for Cures. *Nat. Rev. Chem.* 2, 0148
145. Causal inference in statistics: An overview[Online]. Available: <https://projecteuclid.org/journals/statistics-surveys/volume-3/issue-none/Causal-inference-in-statistics-An-overview/10.1214/09-SS057.full>. [Accessed: 21-July-2025]
146. Perdikaris, P. *et al.* (2017) Nonlinear information fusion algorithms for data-efficient multi-fidelity modelling. *Proc. Math. Phys. Eng. Sci.* 473, 20160751
147. Physics-informed neural networks: A deep learning framework for solving forward and inverse problems involving nonlinear partial differential equations - ScienceDirect[Online]. Available: <https://www.sciencedirect.com/science/article/pii/S0021999118307125>. [Accessed: 21-July-2025]
148. Pearl, J. (2019) The seven tools of causal inference, with reflections on machine learning. *Commun ACM* 62, 54–60
149. Karimi, A.-H. *et al.* (2020) Algorithmic Recourse: from Counterfactual Explanations to InterventionsarXiv
150. Goyal, Y. *et al.* (2019) Counterfactual Visual ExplanationsarXiv
151. Stokes, J.M. *et al.* (2020) A Deep Learning Approach to Antibiotic Discovery. *Cell* 180, 688-702.e13
152. Jin, W. *et al.* (2020) Hierarchical Generation of Molecular Graphs using Structural MotifsarXiv
153. Yang, H. *et al.* (2018) ADMETopt: A Web Server for ADMET Optimization in Drug Design via Scaffold Hopping. *J. Chem. Inf. Model.* 58, 2051–2056
154. Molnar, C. (2020) *Interpretable Machine Learning*, Lulu.com
155. Ramsundar, B. *et al.* (2015) Massively Multitask Networks for Drug DiscoveryarXiv
156. Preuer, K. *et al.* (2019) Interpretable Deep Learning in Drug DiscoveryarXiv

157. Bermejo, M. *et al.* (2020) A Mechanistic Physiologically-Based Biopharmaceutics Modeling (PBBM) Approach to Assess the In Vivo Performance of an Orally Administered Drug Product: From IVIVC to IVIVP. *Pharmaceutics* 12, 74
158. The challenges of explainable AI in biomedical data science | BMC Bioinformatics | Full Text[Online]. Available: <https://bmcbioinformatics.biomedcentral.com/articles/10.1186/s12859-021-04368-1>. [Accessed: 21-July-2025]
159. Liu, L. *et al.* (2021) Prediction of the Blood–Brain Barrier (BBB) Permeability of Chemicals Based on Machine-Learning and Ensemble Methods. *Chem. Res. Toxicol.* 34, 1456–1467

Acknowledgements

The author gratefully acknowledges Yuyang Zhang for his helpful feedback during the final revision of the manuscript.

Competing Interests

The author declares no competing interests.

Data and Code Availability

No original data or code are associated with this review.