Communications in Computational Chemistry
Commun. Comput. Chem., Vol. 7, No. 4, 289–310
doi: 10.4208/cicc.2025.153.01

Regular Article

Application of Modern Intelligent Algorithms in Retrosynthesis Prediction

Jianhan Liao[1], Xiaoxin Shi[2], Ya Gao[3, *], Xingyu Wang[4, *] and Tong Zhu[1,4,5,6, *]

[1]Shanghai Engineering Research Center of Molecular Therapeutics and New Drug Development, School of Chemistry and Molecular Engineering, East China Normal University, Shanghai, 200062, China;

[2]School of Chemistry and Chemical Engineering, Shanghai Jiao Tong University, 800 Dongchuan Road, Minhang, Shanghai, 200240, China;

[3]School of Mathematics, Physics and Statistics, Shanghai University of Engineering Science, Shanghai 201620, China;

[4]NYU-ECNU Center for Computational Chemistry, NYU Shanghai, Shanghai, 200062, China;

[5]Shanghai Innovation Institute, Shanghai, 200003 China;

[6]AI for Science Institute, Beijing, 100080, P.R. China.

[*] Corresponding authors: gaoya@sues.edu.cn; xywang@nyu.edu; tongzhu.work@gmail.com

Received on 2 June 2025; Accepted on 23 June 2025

Abstract: In recent years, the rapid advancements in computer science have spurred the development of various cutting-edge intelligent algorithms. Among these, the transformer, which is built upon a multi-head attention mechanism, is one of the most prominent AI models. The advent of such algorithms has significantly advanced retrosynthesis prediction, though challenges remain in chemical interpretability and real-world deployment. Unlike traditional models, AI-based retrosynthesis prediction systems can automatically extract chemical knowledge from vast datasets to forecast retrosynthesis pathways. This review provides a comprehensive overview of modern intelligent algorithms applied to retrosynthesis prediction, with a particular focus on artificial intelligence techniques. We begin by discussing key deep learning models, then explore available chemical reaction datasets and molecular representations. The discussion extends to the latest state-of-the art in AI-assisted retrosynthesis models, including template-based, template-free, and semi-template-based approaches. Finally, we compare these models across various classifications, highlighting several challenges and limitations of current methods, and suggesting promising directions for future research.

Key words: artificial intelligence, retrosynthesis prediction, machine learning, deep learning.

1. Introduction

Organic synthesis, often regarded as an art due to its reliance on creativity, inspiration and aesthetic judgment [1,2], is a crucial branch of chemistry with extensive applications in drug design and synthetic biology [3–6]. Retrosynthetic analysis, a common method in organic synthesis design [7], involves deducing a synthetic route by working backward from the target compound. The process systematically identifies potential starting materials that can be reassembled to yield the desired compound. However, the growing complexity and diversity of target molecules have made the design of organic synthesis pathways increasingly challenging. To address these obstacles and enhance reproducibility and efficiency, there has been a burgeoning interest in automating organic retrosynthesis [8–10]. Computer-aided synthesis planning (CASP) emerged from this need, tracing its origins to Corey's groundbreaking rule-based synthesis prediction systems, the Logic and Heuristics for Automated Synthesis Analysis (LHASA) [11]. Despite its innovative approach, early rule-based approaches were limited by computational power and data availability.

Recent advancements in computer science have facilitated the development of intelligent algorithms capable of addressing a wide range of tasks [12], such as beam search algorithms, Monte Carlo tree search algorithms, genetic algorithms and neural network algorithms. The influx of big data has propelled the development of numerous artificial intelligence models [13,14], reinvigorating interest in AI applications across chemistry and drug discovery [15,16].

For chemists, CASP poses significant challenges, especially in retrosynthesis prediction, where limited input data can lead to numerous possible outputs. Recent models for retrosynthesis prediction have been designed to automate the identification of candidate reactants from a given product. When these reactants are not commercially available, a recursive expansion strategy is employed, iteratively breaking down the reactants into simpler precursors until all components are accessible or a predefined step limit is reached. Once reliable single-step retrosynthesis is achieved, multi-step retrosynthesis focuses on optimizing the reaction sequence to minimize synthesis steps, costs and waste production.

These models can be broadly categorized into three classes:

1. Template-based models utilize domain knowledge and formal rules derived from prior chemical experiences. Utilizing predefined reaction templates, which specify the transformation of reactants into products, they offer high interpretability and accuracy. However, their applicability is often constrained to scenarios within the predefined knowledge scope, limiting their generalizability.
2. Template-free models, typically devoid of explicit chemical knowledge, such as deep neural networks are considered black-box approaches. While these models have lower interpretability and are prone to violating chemical principles, they have shown promise in discovering new reaction pathways.
3. Semi-template-based models operate in two phases: identifying reaction centers to transform the product into synthons, and then completing these synthons into reactants.

This review provides a comprehensive exploration of contemporary retrosynthetic strategies, focusing on recent advancements in retrosynthesis prediction models. It examines the integration of artificial intelligence (AI) into retrosynthesis prediction, with comparative analysis of commonly utilized data sources and molecular representations. Furthermore, the review evaluates the applications of modern intelligent algorithms across template-based, template-free and semi-template-based models. Finally, it highlights the challenges and outlines potential future directions for this rapidly evolving field.

2. Deep learning algorithms

Artificial intelligence (AI) algorithms are designed to emulate human intelligence, extracting potential rules from datasets and leveraging these rules to make predictions with new data. Deep learning (DL), a rapidly advancing branch of AI, has demonstrated exceptional performance across a variety of tasks, fueled by enhanced computational power and modern algorithms. Generally, DL models can be divided into three categories: supervised learning, unsupervised learning, and reinforcement learning (RL).

In supervised learning, models are trained using datasets containing labeled samples, where the models learn to map input features to output labels. There are two primary types of supervised learning: classification and regression. Classification models predict discrete output labels, while regression models are designed to forecast continuous values. Unsupervised learning, on the other hand, involves training models with datasets of unlabeled samples. These models autonomously identify patterns and relationships in the data without explicit guidance. Reinforcement learning allows an agent to learn through trial and error within a specific environment. The agent receives rewards for actions that produce desirable outcomes and penalties for undesirable ones, aiming to develop a strategy that maximizes expected rewards over time. Most retrosynthesis prediction models employ the supervised learning strategy, as illustrated in **Figure 1**.

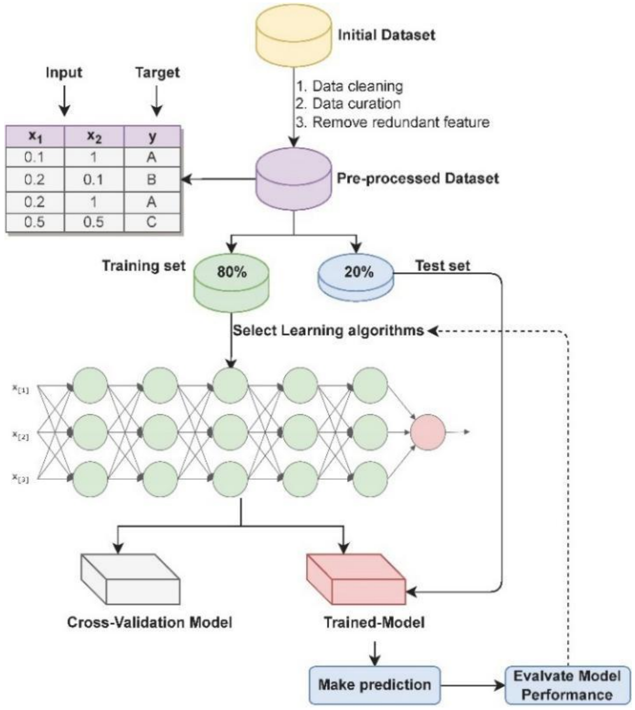

Figure 1. The process of supervised learning method.

Common DL algorithms used in retrosynthesis prediction include sequence to sequence (Seq2Seq) models, graph neural networks (GNNs), reinforcement learning (RL), and various search algorithms.

2.1 Sequence generation model

Molecules can be represented as SMILES-based sequences, allowing the retrosynthesis prediction task to be transformed into a sequence-to-sequence task. Widely applied in the *natural language processing (NLP)* domain, Seq2Seq models have demonstrated their efficacy in chemical sequence modeling. These models generate sequences of chemical reactants from a chemical product, even when input and output sequences vary in length. In this review, we emphasize recurrent neural network-based Seq2Seq models and attention-based Seq2Seq models.

To address sequence generation challenges, such as those found in machine translation, recurrent neural networks (RNNs) were introduced for encoding and decoding tasks [17,18]. Unlike feed-forward neural networks, RNNs employ hidden states to store previous information. The RNN encoder translates an input sentence into a fixed-length vector, while the decoder sequentially generates the corresponding target words. The framework of RNN is illustrated in **Figure 2**. However, traditional RNN models face difficulties in capturing long-distance dependencies and lack the ability to parallelize computations. The Bidirectional Long Short-Term Memory (biLSTM), a variant of RNN, overcomes these limitations by selectively retaining long-distance dependencies using gating mechanisms. Additionally, the attention mechanism, which prioritizes limited resources on crucial information, enhances the biLSTM framework by allowing hidden states to incorporate global information, addressing the issue of non-parallelizable computations [19]. To further improve global focus, a multi-step attention mechanism is introduced at each decoder layer. The Transformer model, pioneered by Vaswani et al., features encoders and decoders that depend solely on Multi-Head Self-Attention mechanisms. This design enables effective capture of long-range correlations within sequences. In recent years, transformer-based models have emerged as a leading approach in purely data-driven retrosynthesis prediction, thanks to their exceptional performance.

The attention mechanism in deep learning assigns weights to different parts of the input data, prioritizing areas of greater importance by assigning higher weights. The self-attention mechanism, a specific type of attention, applies this concept within the same sequence, enabling the model to discern relationship between any two positions. This is essential for understanding the structural and contextual information embedded in the sequence. The self-attention mechanism is implemented using query, key, and value vectors: the query vector assesses the similarity with key vectors, and the resulting weights are applied to the value vectors to compute the output.

More specifically, for a set of column vectors, $H = [h_1, \dots, h_T] \in R^{D_h \times T}$, the self-attention mechanisms can be conceptualized as a process that establishes interactions between different vectors h_i in a linearly projected space. The encoding formulas of self-attention mechanisms are as follows:

$$\text{self-att}(Q, K, V) = \mathrm{V} \cdot \operatorname{softmax}\left(\frac{K^{\top} Q}{\sqrt{D_k}}\right) \tag{1}$$

$$Q = W_q H,\ K = W_k H,\ V = W_v H \tag{2}$$

Here, D_k denotes the dimension of the column vectors in the input matrices Q (queries) and K (keys). D_v denotes the dimension of column vectors in the matrices V (values). $W_q \in R^{D_k \times D_h}$, $W_k \in R^{D_k \times D_h}$, $W_v \in R^{D_v \times D_h}$ are three projection matrices.

$$\text{MultiHead}(H) = W_o[\text{head}_1; \cdots; \text{head}_m] \tag{3}$$

$$\text{head}_m = \text{self-att}(Q_m, K_m, V_m) \tag{4}$$

$$\forall m \in \{1, \cdots, M\}, \quad Q_m = W_q^m H,\ K = W_k^m H,\ V = W_v^m H \tag{5}$$

The use of Multi-Head Self-Attention allows for the capture of varied interaction information across multiple distinct projection spaces. When the self-attention model is applied within M such projection spaces, it can be mathematically represented as follows:

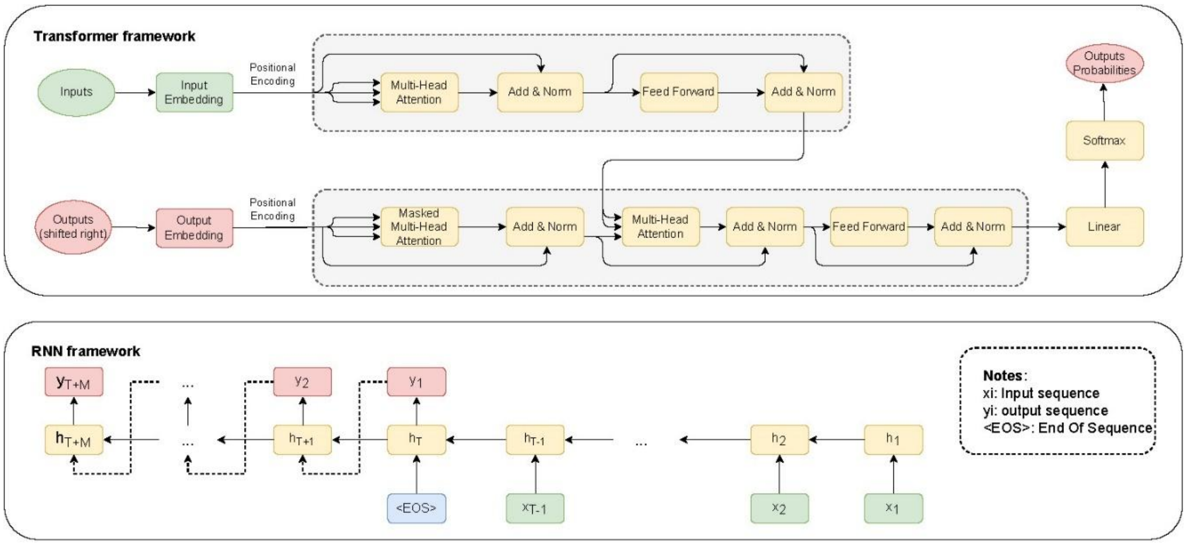

Figure 2. Transformer framework and RNN framework.

Here $W_o \in R^{D_h \times M_{d_v}}$ is the output projection matrix, and $W_q^m \in R^{D_k \times D_h}$, $W_k^m \in R^{D_k \times D_h}$ and $W_v^m \in R^{D_v \times D_h}$ are the projection matrices for $m \in \{1, \ldots, M\}$.

Figure 2 illustrates the network architecture of the Transformer model, which can be divided into two parts: the encoder and the decoder. The encoder comprises multiple layers of multi-head attention modules. The decoder generates the target sequence autoregressively, consisting of masked self-attention modules, decoder-to-encoder attention modules, and feedforward neural networks.

Besides RNN-based models and attention mechanisms-based models, Gehring et al. proposed a framework for sequence modeling [20] known as the convolutional sequence to sequence (ConvS2S) model. Its encoders and decoders consist of multilayer convolution neural networks, which are more efficient than RNNs in some cases.

2.2 Graph neural networks

Molecules can be represented not only by encoding them into sequences but also as undirected weighted graphs, a data structure derived from graph theory. These consists of a set of vertices, a set of edges, and a set of global information. Each edge is assigned a weight, and connections between vertices are directionless. For a detailed discussion about graph-based molecular representations, please refer to Section 4.3.

There are three general types of prediction tasks on graphs: graph-level, node-level, and edge-level. The prediction of chemical molecules typically falls under the graph-level category, which can be solved with Graph Neural Networks (GNNs). GNNs are promising parameter-efficient tools for learning the structural information of graphs, enabling the predictions of molecular transformations in reactions [21]. A GNN is an optimizable transformation applied to all attributes of the graph that preserves graph symmetries (permutation invariances). Sperduti was the pioneer in applying neural networks to directed acyclic graphs [22]. This approach is also applicable to the undirected graph representation of chemical molecules.

Figure 3 illustrates an example of a GNN using the “message passing neural network” framework for a binary classification task, which can easily be extended to the multi-class or regression tasks. With the numerical representation of graphs as input, this GNN learns new embeddings for all graph attributes (nodes, edges, global) without using the connectivity of the graph. This GNN uses a separate multilayer perceptron (MLP) on each component of a graph referred to as a GNN layer. For each graph attribute vector, the MLP is applied, and a learned vector is generated. Finally, it makes predictions by pooling information, such as gathering information from edges to nodes.

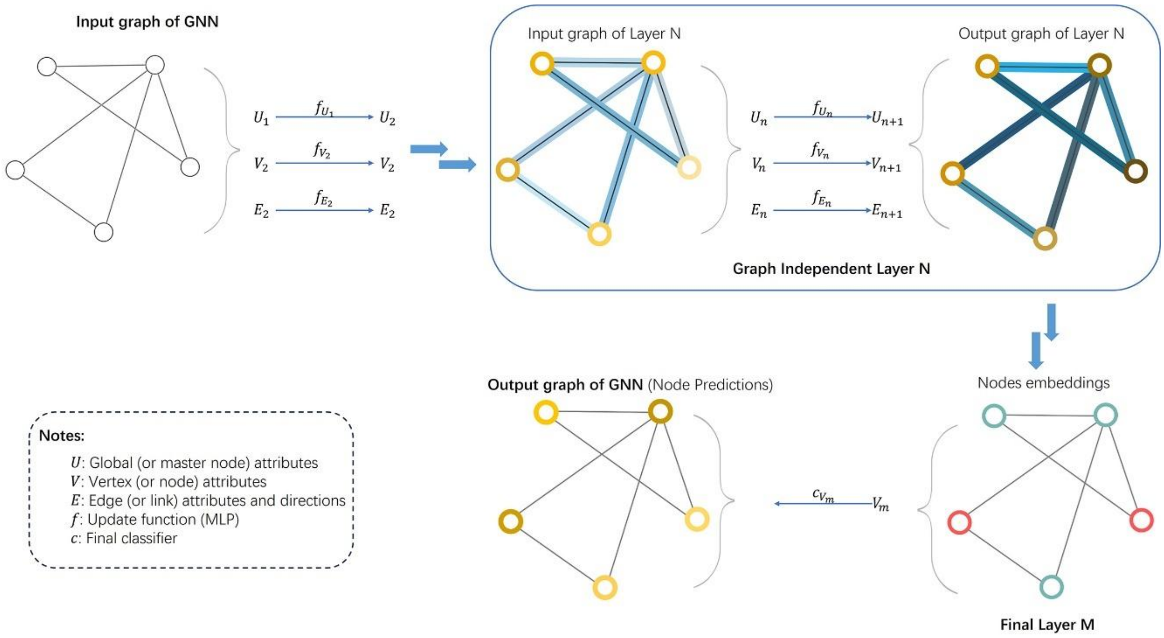

Figure 3. GNN framework with message passing neural network for classification tasks.

Researchers have further proposed Recurrent Graph Neural Networks (RecGNNs) [23,24], where neighbor information is propagated iteratively to update the representations of target nodes. Due to the significant success of convolutional neural networks (CNNs) in computer vision, convolutional operations were introduced into GNNs, leading to the development of Graph Convolutional Networks (GCNs) [25]. The convolution operation in GCN is a weighted average of the features of the graph to aggregate information about features and their neighbors. However, the weights generated by the aggregation operation lack permutation invariances. To overcome this problem, researchers have introduced attention mechanisms into GNN and proposed Graph Attention Networks (GATs) [26] and Gated Attention Networks (GAANs) [27]. Building upon these advances, Graph Autoencoders (GAEs) [28], Graph Generation Networks (GGNs) [29], and Spatio-Temporal Graph Convolutional Networks (STGCNs) [30] have been further developed.

2.3 Reinforcement learning

Reinforcement Learning (RL) is a type of unsupervised learning method [31]. It addresses problems where an agent learns by interacting with the environment to achieve specific objectives, such as maximizing rewards. Similar to deep learning, a crucial challenge in RL is the allocation of contributions. Each action does not receive direct supervised feedback but relies on a delayed supervised signal (reward) from the entire model. The primary distinction between RL and supervised learning is that RL does not require a predefined "correct" strategy as supervised information; instead, it focuses on maximizing expected returns by refining strategies based on delayed rewards.

In RL, two interacting entities exist: the agent and the environment. The agent perceives the state and reward of the environment, engaging in learning and decision-making. Decision-making involves selecting actions based on the current state of the environment, while learning adjusts strategies in response to received rewards. The environment comprises all external elements to the agent, subject to changes in its state due to the agent's actions and providing corresponding rewards to guide the agent's behavior.

The fundamental components of RL include:

(1) States (s): A description of the environment, which can be discrete or continuous, forming the state space S.

(2) Actions a: Representing the agent's behavior, also in discrete or continuous forms, forming the action space A.

(3) Policy $\pi(a|s)$: A mapping that determines the agent's next action a based on the environment's state s.

(4) State transition probabilities $p(s^{'}|s,a)$: The probability of the environment transitioning to a new state s' after the agent takes action a from the current state s.

(5) Immediate rewards $r(s,a,s^{'})$: Scalar functions that provide feedback to the agent based on its action a in the current state s, often correlated with the subsequent state s'.

The objective of RL is to learn a policy $\pi(a|s)$ that maximizes the expected return. The objective function is represented as:

$$\mathcal{J}(\theta) = \mathbb{E}_{\tau \sim p_\theta(\tau)}[G(\tau)] = \mathbb{E}_{\tau \sim p_\theta(\tau)}\left[\sum_{t=0}^{T-1} \gamma^t r_{t+1}\right]$$

Here, θ represents the parameters of the policy function. Value functions are used to evaluate the expected return of a policy π, including state value functions and state-action value functions (Q-Function). The policy can be iteratively optimized based on these value functions. Additionally, the expected return can be maximized by directly searching the policy space, using methods like gradient-based optimization [32,33] and gradient-free optimization.

Deep Reinforcement Learning combines RL and deep learning methodologies, employing RL to define problems and optimization goals, using deep learning to address the modeling of policy and value functions, and subsequently employing error backpropagation algorithms to optimize the objective function. Mnih et al. proposed Deep Q-Networks (DQNs) [34], a cornerstone in the field of deep RL. DQNs use convolutional neural networks to estimate Q values. Two pivotal measures in the Deep Q-Network are:

1. Freezing of target networks: The parameters of the target network are fixed for a specified duration to stabilize learning objectives.

2. Experience replay: An experience pool is constructed to eliminate data correlations. This pool stores recent experiences gathered by the agent as a dataset. During training, random samples are drawn from the experience pool to replace current samples. This approach breaks the similarity between adjacent training samples, preventing the model from converging to local optima. The learning process of DQNs is illustrated below.

Algorithm: DQN with Experience Replay

```
Input: State space $S$, Action space $A$, Discount rate $\gamma$, Learning rate $\alpha$
Initialize experience pool $D$ with capacity $N$;
Randomly initialize parameters of the Q network $\phi$;
Randomly initialize parameters of the target Q network $\hat{\phi} = \phi$;
Repeat
    Initialize the starting state $s$;
    Repeat
        In state $s$, select action $a = \pi^{\epsilon}$;
        Execute action $a$, obtain immediate reward $r$ and the new state $s^{'}$;
        Place $s, a, r, s^{'}$ into $D$;
        Sample $ss, aa, rr, ss^{'}$ from $D$;
        $y = \begin{cases} rr, & ss^{'} \text{ is terminal state} \\ rr + \gamma \max_{a} Q_{\hat{\phi}}(ss^{'}, a^{'}), & \text{otherwise} \end{cases}$
        Train the Q network with the loss function: $\left(y - Q_{\phi}(ss, aa)\right)^2$;
        $s \leftarrow s'$;
        Every C steps, execute action: $\hat{\phi} \longleftarrow \phi$;
    Until $s$ is the terminal state;
Until $\forall s$ and $a$, $Q_{\phi}(s,a)$ converges;
Output: $Q_{\phi}(s,a)$
```

Since the advent of value-based approaches, numerous extensions have been introduced to enhance their efficiency and applicability [35]. In addition, model-based methodologies have been proposed [36], enabling the prediction of post-action states through predictive models and direct optimization of policy networks. Deep RL also addresses more complex decision-making problems, such as those with goal conditions [37], hierarchical task decomposition [38], and multiple agents [39]. It has achieved significant success across diverse applications, ranging from games [40], robotics [41], and autonomous driving [42] to molecule generation [43]. Such advancements are widely recognized as crucial milestones toward the realization of general AI [44].

2.4 Search algorithms

Search algorithms retrieve stored information from a data structure or compute it within a search space, forming the foundation for multi-step retrosynthesis prediction in planning synthesis routes. Generally, these algorithms are classified into two categories: uninformed searches and informed searches. Uninformed searches do not use information regarding the cost of state transitions; common examples include depth-first search and breadth-first search. In contrast, informed searches use heuristic functions to estimate the distance between the current stage and the goal, thereby guiding the search process. While not necessarily optimal, this approach provides favorable solutions within a reasonable time frame. Best-first searches represent typical heuristic searches employing a priority queue concept. The OPEN list contains currently traversable nodes, while the CLOSED list stores traversed nodes. Beam search enhances best-first search by expanding the most promising nodes within a limited set [45]. A search amalgamates the merits of uniform cost search and best-first search, ensuring optimality in solutions [46]. In this context, the cost of each state comprises the actual cost from the starting state to the current one and the heuristic cost from the current state to the goal state. Monte Carlo Tree Search (MCTS) [47] refines value estimates from the current state to the goal state. AlphaGo [48] stands as one of the most renowned applications of MCTS, where it explores potential moves and tracks outcomes within a Go search tree. MCTS consists of four phases: Selection, Expansion, Simulation and Backpropagation. (see **Figure 4**)

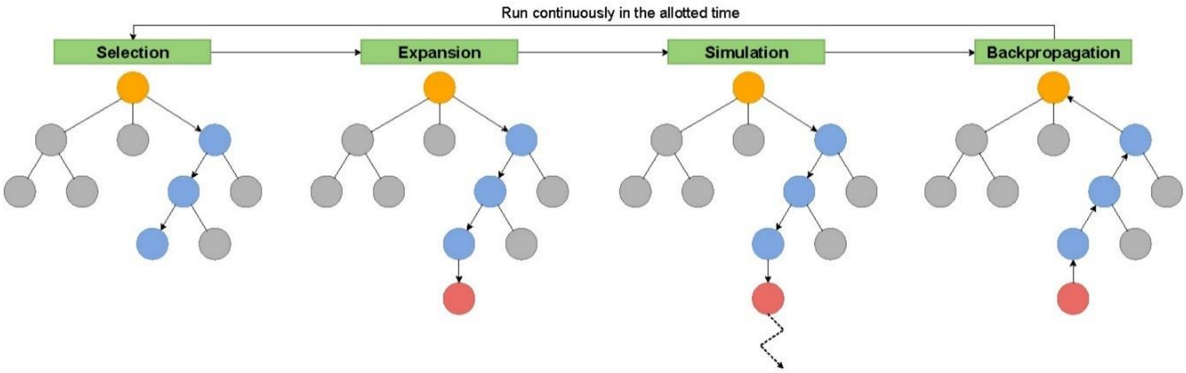

Figure 4. The process of MCTS: selection, expansion, simulation, and backpropagation.

3. Data sources

In CASP tasks, whether employing symbolic AI or purely data-driven modeling, a dataset that can be parsed by a computer is a fundamental requirement. The quality of the dataset determines the model's upper performance limit. It is no exaggeration to state that the quality of the dataset is often more important than the model itself [49]. Therefore, computational chemists must pay particular attention to the characteristics of the input dataset. This section provides a summary and comparison of commonly used chemical reaction databases.

Journals and publishing houses have made their datasets available under licensing agreements in computer-readable format, through automatic extraction by algorithms and expert manual coding. Examples include the Reaxys database, published by Elsevier, which， as of 2023, contains over 73 million reactions. It offers comprehensive and up-to-date journal and patent coverage from 16,000 journals and 105 patent offices. To extract information from chemical patents, Elsevier and the University of Melbourne, Australia initiated a project based on NLP models, called ChEMU [50]. The Chemical Abstracts Service (CAS) encompasses approximately 150 million reactions spanning from 1840 to 2023, including organic, inorganic, total synthesis of natural products, and biotransformation reactions, making it the largest provider of reaction data. Its data sources derive from journals, patents, dissertations, and seminal reference works. Furthermore, smaller-scale datasets include SPRESI, developed by InfoChem, which encompasses 4.6 million reactions spanning the period from 1974 to 2014. Another prominent dataset, Pistachio, created by NextMove Software, comprises patent data from 1976 to 2023, encompassing a vast corpus of over 13,118,970 reactions. Among researchers, the most extensively employed dataset is a subset of patent data extracted by Lowe during the period from 1976 to 2016, which encompasses 3.3 million reactions. This dataset is presently the sole publicly accessible repository of reaction data and is commonly called USPTO [51]. Moreover, USPTO 50K, a subset and preprocessed iteration of Chemical reactions from USPTO, is composed of 50,000 randomly selected reactions, covering ten distinct reaction types [52]. USPTO-MIT [53] is also a commonly used subset, which contains a wider range of reagents and possible catalysts compared to USPTO-50K. The specific details of commonly used datasets are listed in **Table 1**.

Although the datasets mentioned above include details on molecular structures, reaction conditions (e.g., solvents, catalysts, reagents), and yields, they are not immune to errors. Additionally, the prevalence of positive data in most patents and literature contributes to an uneven distribution of product representations [54,55], which can negatively impact model performance. Furthermore, within the CASP framework, instances of failed reactions play an important role, especially in situations concerning regioselectivity and chemoselectivity. To address these challenges, efforts have been made to publish data that ensures greater consistency [56]. IBM has developed a method leveraging Natural Language Processing (NLP) to extract experimental procedures from patents and scientific literature, thereby creating structured, automation-friendly formats [57]. The Pistoia Alliance has collaborated with Elsevier to define a Unified Data Model (UDM) for the exchange of reaction information. Electronic laboratory notebooks (ELNs), a novel dataset extracted from the electronic laboratory notebooks of a large pharmaceutical company, are not subject to the publication bias toward high-yielding reactions [58,59].

Of notable mention is that, when comparing various data sources, such as patents (USPTO and Pistachio), literature and patents (Reaxys), and industrial data (AstraZeneca ELN), despite similarities in the size of their template sets, they differ in the coverage of reaction space33. Reaxys stands out for its extensive and uniquely diverse collection of reaction templates, providing a broader reaction space [60].

Table 1. Overview of dataset used for retrosynthesis prediction models.

Dataset	Source	Sample size	Reaction space coverage	License/Accessibility Notes
Reaxys	journals and patent	7300k	+++++	Commercial annual license, includes confidentiality clauses
ChEMU	patents	N/A	N/A	Creative Commons Attribution 4.0 International (CC BY 4.0)
CAS	journals and patent	15000k	N/A	Commercial license, contact CAS for public or commercial use
SPRESI	literature	4600k	N/A	Commercial database, typically accessed via institutional/national site licenses
Pistachio	USPTO + EPO	9000k	++	Commercial license (NextMove Software product), contact directly for details
USPTO-full	USPTO	3300k	++++	Public domain within U.S., USPTO reserves international copyright, may contain third-party copyrighted content
USPTO 50K	USPTO	50k	+++	Derived from public domain USPTO patents, no explicit license, academic attribution generally expected

Note: + symbols indicate reaction space coverage level (+ = low, +++++ = very high)

N/A: Not Available / Not Applicable, indicating the information is not provided by the dataset source or is outside its scope.

4. Molecular representation

In CASP tasks, the quality of the dataset and the art of feature engineering play determining roles in the performance of the model. Therefore, chemists have developed numerous distinct molecular representation methods with the aid of mathematical tools. These methods aim to encapsulate the complete information of molecules through abstract mathematical symbols. 1D molecular representation methods can solely represent global molecular properties exclusive of structural patterns, such as pKa, logP, etc. 2D molecular representation methods can represent structural patterns without explicit 3D information, including SMILES (Simplified Molecular Input Line Entry System) [61,62], fingerprints, and molecular graphs, which are the mainstream methods used in retrosynthesis tasks. 3D molecular representation methods, such as image-based methods, can contain high-dimensional information. However, higher dimensionality does not always guarantee better performance. In recent years, a 3D molecular representation learning framework has been proposed to automatically capture richer information in higher dimensions [63].

4.1 Molecular string representation

SMILES is the most widely adopted molecular string representation system for encoding molecular structures. It combines specific syntax rules with chemical principles to represent molecular structures in a rigorous and compact form. One of the advantages of SMILES is its ability to transform reaction prediction tasks into machine translation tasks. For sequence modeling problems, leveraging natural language processing (NLP) models in the field of artificial intelligence can solve them efficiently [64]. For example, the Transformer architecture based on the attention mechanism is one of the most favored NLP models among computational chemists. For SMILES representation of chemical reactions, reactants, reagents, and products can be linked together using symbols, similar to the molecular fingerprint method. The ">" symbol is used to indicate the direction of the reaction, as in the format: "Reactants > Reagents > Products". However, SMILES grammar is sequence-sensitive and has trouble dealing with stereochemical representation. SMARTS, an extension of the SMILES language, is designed for describing molecular patterns and properties. SMARTS can be used to create queries. One notable feature of SMARTS is its allowance for the use of wildcards to represent atoms and chemical bonds. This flexibility has made SMARTS a widely used tool for efficient and versatile chemical structure searching.

Self-Referential Embedded Strings (SELFIES) [65] is a method for representing molecular structures that is both 100% robust and human-readable. SELFIES was proposed to address the limitation of SMILES, such as grammar violations. InChI [66], another string-based representation, offers the advantages of uniqueness and reversibility compared to SMILES. Unlike SMILES, these methods no longer involve atom-atom mapping to identify reaction centers. The SMILES representation of caffeine is shown in the following **Figure 5**, including the process of ensuring its SMILES representation.

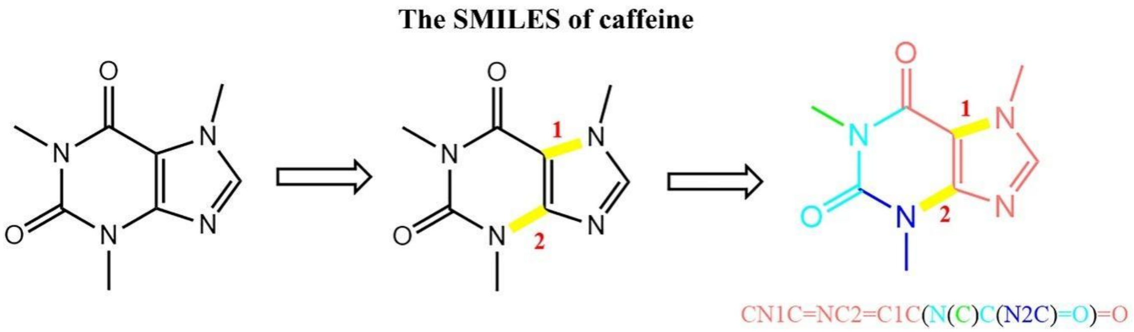

Figure 5. The process of getting the SMILES representation of caffeine.

4.2 Molecular fingerprints

Molecular fingerprints are another valuable tool in cheminformatics for representing molecules. The core idea behind molecular fingerprints is to map a molecular into a bit string or a numeric array of length l, where each bit encodes whether the molecule contains a specific substructure feature. Molecular fingerprints offer several advantages, including high computational efficiency and ease of retrieval, making them an ideal choice for molecular similarity assessments. Common approaches include substructure key-based fingerprints, path-based fingerprints, and circular fingerprints. This review focuses on widely used molecular fingerprint methods. For a comprehensive introduction to molecular fingerprints and related software, please refer to Cereto-Massagué's work [67].

Substructure key-based fingerprints represent compounds as bit strings, encoding the presence of specific substructures or features from a predefined list of structural keys. The MACCS fingerprint system [68] offers two options: a 960 bits variant and a more compact 166-bit version, both based on SMARTS patterns of structural keys. Despite its reduced size, the 166-bit variant effectively captures most chemically relevant features, making it suitable for tasks such as drug discovery and virtual screening. Similarly, the PubChem fingerprint [69] comprises 881 structural keys, providing a comprehensive representation of diverse substructure features and forming a basis for similarity searches in the PubChem database. The BCI fingerprint [70], with user-customizable options and a standard substructure dictionary of 1052 keys, offers flexibility in its generation [71]. Furthermore, the TGD and TGT fingerprints [71,72], calculated from 2D molecular graphs, present two-point and three-point pharmacophore representations, consisting of 735 and 13,824 bits, respectively. These fingerprints, with their distinct characteristics, cater to a wide range of cheminformatics applications, allowing researchers to effectively explore and analyze chemical compound data.

Path-based fingerprinting operates by scrutinizing all molecular fragments that follow predefined paths (typically linear), up to a specific number of bonds. Subsequently, each of these paths undergoes hashing to generate a unique fingerprint. These fingerprints are widely used for substructure searches and molecular filtering due to their computational efficiency. Among them, the "Daylight Fingerprint" stands out prominently [73], comprising as many as 2048 bits that meticulously encode all possible connectivity paths within a molecule up to a defined length.

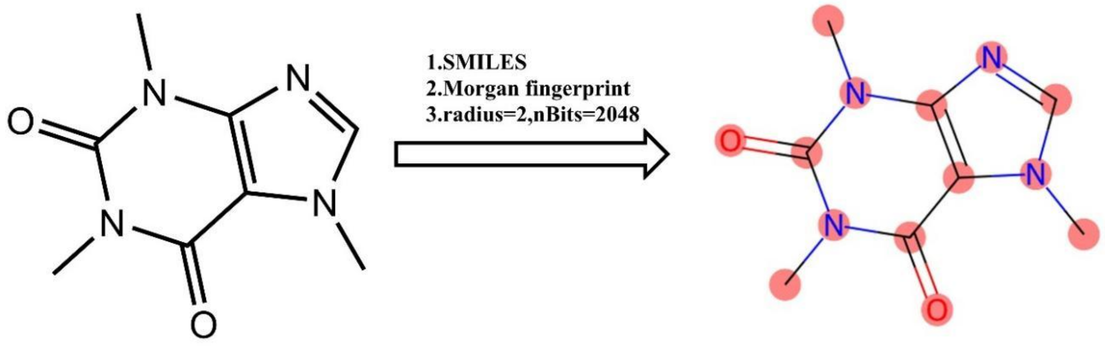

Figure 6. The result of the caffeine structure with highlighted atoms which are related to Morgan fingerprint.

Circular fingerprints focus on recording the environment surrounding each atom within a defined radius. While they are less suitable for substructure verification queries due to the possibility of identical fragments exhibiting distinct environments, they are valuable for full-structure similarity searches. Molprint2D encodes the atomic environments of each atom in a molecule's connectivity table, representing these environments as strings of varying sizes [74,75]. ECFP (Extended-Connectivity Fingerprints) is an extension of the circular fingerprint based on the Morgan algorithm [76]. They represent cyclic atom neighborhoods and generate variable-length fingerprints. The commonly used ECFP variant has a diameter of 4, often referred to as ECFP4. A diameter of 6 (ECFP6) is also quite common. FCFP (Functional-Class Fingerprints) is a variant of ECFP, indexing the function roles of atoms. Different atoms with similar functions are not distinguished in the fingerprint. It can represent stereochemistry information which can further be used to infer structure-activity relationships. **Figure 6** shows how to convert the caffeine structure from the SMILES representation to Morgan fingerprint.

4.3 Molecular graphs

With the rapid advancements in graph neural networks, molecular graphs have gained significant attention from researchers in the CASP field. An undirected graph serves as a foundational data structure in graph theory, consisting of nodes and edges with associated weights. In such graphs, edges lack explicit directionality, allowing for bidirectional connections between nodes A and B. Representing a finite graph, the adjacency matrix is a square matrix where each element denotes whether nodes are interconnected by edges. The matrix's size corresponds to the number of vertices in the graph, and its diagonal elements are invariably zero. A value of 1 at the intersection of row i and column j indicates a direct edge between node i and node j. However, the adjacent matrix exhibits a space complexity of $O(n^2)$, where n is the number of vertices in the graph, making it inefficient for large graphs. To enhance computational efficiency, especially with larger matrices, adjacency matrices can be transformed into eigenvectors representing nodes, edges and global properties. These eigenvectors are often used as input features for graph neural networks.

Molecules, the elemental constituents of matter, consist of atoms and electrons arranged in three-dimensional space. Despite interactions among all particles, a stable separation between two atoms forms a covalent bond. Varied atomic pairs and bonding configurations, such as single and double bonds, manifest distinct interatomic distances. This inherent characteristic facilitates the representation of molecules as graphs, where atoms serve as nodes and chemical bonds as edges [77–79]. **Figure 7** illustrates the graph representation of Caffeine, showcasing its molecule structure, molecule graph and corresponding adjacency matrix.

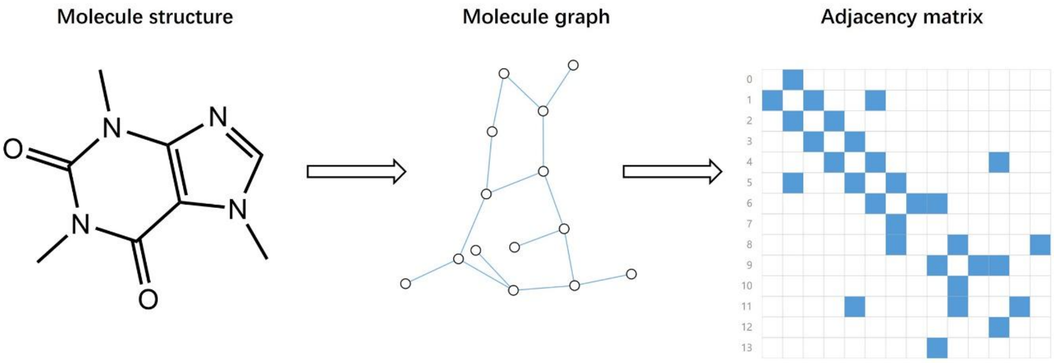

Figure 7. The graph representation of Caffeine.

Compared to SMILES and molecular fingerprints, molecular graphs offer a richer representation of chemical structures. They capture detailed information about atom types, bond types, topologies, as well as 3D information such as bond lengths and bond angles. Additionally, molecular graphs are invariant to atom ordering, making them robust for various applications. However, the practical application of molecular graphs relies on efficient algorithms to extract these representations from molecular structures [78,80].

For reaction graphs, extracting reactions using pre-trained models presents a promising approach. Atom mapping further enhances the effectiveness of reaction graphs by enabling a single condensed reaction graph (CGR) to represent chemical reactions [81]. This CGR is a superposition of reactant and product graphs, providing a comprehensive view of the reaction dynamics.

5. Retrosynthesis strategy evaluation

5.1 Candidate reaction evaluation

In retrosynthesis, “combinatorial explosions” present a significant challenge. Scientists aim to constrain recursive unfolding to the most promising bond breaks, thereby focusing on structures that are more easily synthesized.

The synthesizability of molecular structures is crucial in evaluating candidate reactions. The Synthesis Accessibility Score (SA Score) measures this by evaluating the contribution of fragments, which scale linearly with commonly synthesizable structural features, while penalizing the presence of rare and complex structural features [82,83]. Chematica introduces a metric for assessing synthetic difficulty by considering structural complexity, reaction step length, reaction conflicts and protecting groups. The SCScore is based on the principle that reaction products should exhibit higher synthetic complexity than their reactants [84,85]. Other evaluation methods include the support vector machine-based DRSVM [86] and various current complexity metrics [87].

5.2 Model evolution

In a CASP modeling workflow, model evaluation plays a pivotal role. CASP tasks, due to their specificity, differ significantly from conventional regression and pattern recognition tasks. To select models suitable for practical retrosynthesis tasks, evaluation metrics tailored to these tasks must be employed. Retrosynthesis tasks are generally divided into two categories: single-step retrosynthesis and multi-step retrosynthesis prediction.

For single-step retrosynthesis, Top-N accuracy calculation is a commonly used metric to evaluate the performance. It examines whether the entire set of ground truth precursors-the actual reactants reported in the template library for the corresponding target molecule-is included among the first N precursors suggested by the model. This metric demands an exact match in molecular structure, which can be measured using a molecular similarity. A similarity score of 1 denotes identical structures [88]. Additionally, some alternative evaluation metrics for single-step retrosynthesis have been introduced [89]. For multi-step retrosynthesis, evaluation can be achieved by using single-step retrosynthesis methodology repeatedly.

Beyond Top-N accuracy, comprehensive evaluation of retrosynthesis models requires additional metrics that assess practical utility. Chemical validity measures whether predicted reactants can realistically undergo the proposed transformation, ensuring adherence to fundamental chemical principles - a particular concern for template-free models that may generate chemically implausible suggestions. Synthetic accessibility evaluates the practical feasibility of obtaining predicted starting materials, with metrics like SA Score and SCScore helping prioritize routes leading to readily available compounds. Performance on rare reaction types is equally important, as most datasets exhibit significant class imbalance favoring common transformations, making evaluation on underrepresented reaction classes essential for assessing true generalization capabilities. Additional considerations include prediction diversity, computational efficiency, and ultimately, experimental validation rates, though the latter remains limited due to the cost and time requirements of laboratory verification.

6. Template-based models

Template-based models often involve matching the target molecule with an entire template library. The subgraph isomorphism problem is then solved to obtain candidate reactants. The core of template-based systems lies in the use of retrosynthesis templates. As shown in **Figure 8**, a reaction template is represented by molecular subgraph patterns that encode changes in the connectivity of atoms during a reaction. Mathematically, a retrosynthesis template T is denoted as the following rule:

$$T: p^T \rightarrow \{r_i^T\}_{i=1}^{n_r}$$

Where p^T is a subgraph of the product P and can be regarded as the reaction center, while r_i^T is the subgraph of the ith reactant.

Starting with a target molecule, a template is selected following predefined rules and is applied to the target molecule to determine the reactants. While template-based methods offer better interpretability and accuracy than template-free methods, they are computationally demanding and have limited generalization outside the template library. The mission of modern intelligent algorithms is to lower the computational complexity of this process.

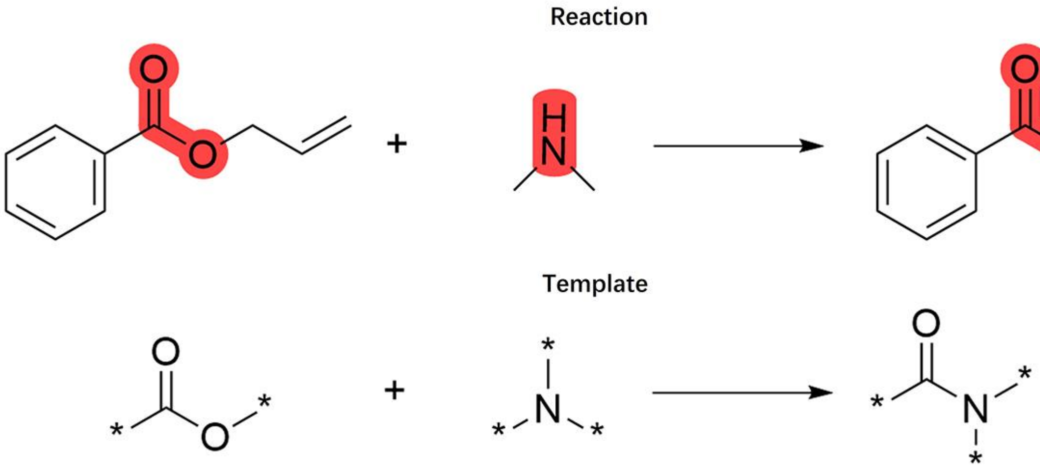

Figure 8. Illustration of a chemical reaction and its retrosynthesis template.

Traditionally, reaction rules have been defined and hand-coded by experts. Szymkuc et al. provided a review on using reaction templates coded by human experts for synthetic planning [90]. With the reaction space growing exponentially at a rate of 4.4% per year [91], manual coding becomes an overwhelming task. An alternative approach to reaction coding utilizes algorithms that extract reaction centers via atom-to-atom mapping to identify correspondences between reactants and products [92–95]. For a given reaction, one can identify the set of atoms that change bond connectivity as reaction centers. Then, the reaction centers and adjacent atoms are algorithmically extracted and generalized to form the corresponding retrosynthesis template.

With reaction templates available, Coley et al. proposed a retrosynthesis method based on molecular similarity metrics [96], such as Morgan2noFeat, Dice similarity, Tanimoto similarity and Tversky similarity. This approach decomposes target molecules solely based on analogy to known reaction precedents, inherently disfavoring creative disconnections. Segler et al. used extended-connectivity fingerprints (ECFP) as input and constructed a deep neural network-based model that can learn to resolve reactivity conflicts and prioritize the most appropriate transformation rules, which is one of the first ML-based template models [97]. This model addresses the multi-class classification problem of categorizing similar templates into subgroups and its performance is often used as a benchmark in template-based approaches. Watson et al. proposed a template-based approach using reverse reaction transforms (RRTs) [98]. RRTs are extracted from clusters of similar reactions. By searching possible synthesis routes in the RRT-repository, this method decomposes a target molecular into fundamental building blocks. Genheden et al. developed the retrosynthesis software AiZynthFinder [99], which is based on a Monte Carlo tree search that recursively disconnects molecules into purchasable precursors. The tree search is guided by an artificial neural network strategy that suggests possible precursors by utilizing a library of reaction templates. Park et al. proposed undersampling based on the similarity (random, dissimilarity) clustering of molecular structures of products to address the class imbalance problem in chemical reaction datasets [100], significantly improving prediction accuracy. Chen et al. proposed a local retrosynthesis framework, LocalRetro [101], which assums that molecular changes occur mostly locally in the process of reaction, complemented by a global attention mechanism to account for the nonlocal effects. Seidl et al. proposed a template-based single-step retrosynthesis model based on modern Hopfield networks [102], which learns the encoding of molecules and reaction templates to predict the correlation of the template with a given molecule. This template representation allows generalization across different reactions. AiZynthTrain, developed by Genheden et al [103], is a robust, reproducible, and extensible end-to-end retrosynthesis model. Its process includes two pipelines that build a template-based one-step retrosynthesis model and a ring breaker model, highlighting the important role of heuristics. Dai et al. proposed a conditional graph logic network model based on a hierarchical sampling approach [104]. A conditional graph logic network is built on graph neural networks that learn when the rules in a reaction template should be applied, implicitly considering whether the final reaction is chemically feasible and strategic. RetroComposer [105], proposed by Yan et al, can synthesize new templates in addition to using training templates. They developed an effective candidate scoring model that can capture atomic level transformation.

In a broad sense, template-based models can include quantum-computation-based retrosynthesis models, as quantum computation can generate new reaction templates. Liu et al established a reaction kinetics-based retrosynthesis planning framework to design synthetic pathways [73]. The forward analysis component consists of transition state theory (TST)-based reaction kinetic model and DFT calculations. The retrosynthesis planning component includes Decision tree model and breadth-first search algorithm. To address the issue of poor sample quality in datasets, Toniato et al. proposed supplementing missing data for model retraining using first-principal computation [106].

Table 2. Overview of retrosynthesis prediction performance for template-based methods.

Methods	**Algorithm**	**Dataset**	**Features**	**TOP-1**	**TOP-5**	**TOP-1**	**TOP-5**	**source code availability**
				with reaction class		**without reaction class**		
Retrosim	Similarity	-	fingerprint	52.9	81.2	37.3	63.3	Y
Park	Taylor－Butina algorithm	Reaxys	SMILES+fingerprint	-	-	51	84	Y
LocalRetro	Attention mechanism	USPTO-50k	graph	63.9	92.4	53.4	85.9	Y
Seidl et al.	Hopfield Networks	USPTO-50k	SMILES	-	-	51.8	81.2	Y
Neuralsym	ANN	Reaxys	ECFP	55.3	81.4	44.4	72.4	Y
GLN	GLN	USPTO-50k	graph	64.2	85.2	52.5	75.6	Y
RetroComposer	Multiple DL models	USPTO-50k	graph	65.9	89.5	54.5	83.2	Y

7. Template-free models

Recently, template-free methods have gained popularity in the field of retrosynthesis as they circumvent the computationally demanding task of subgraph matching. These approaches leverage textual representations of molecules, such as SMILES or InChI codes, transforming retrosynthesis into a translation task. This paradigm shift enables the use of advanced deep learning techniques, eliminating the need for direct atom-to-atom mapping to identify reaction centers. Predominantly data-driven, these approaches generally do not require the integration of explicit chemical knowledge. When extensive and relevant data are available, these methods can achieve satisfactory results. The following provides an overview of these approaches, which are categorized into four groups: deep neural networks, sequence-to-sequence models, graphical neural networks, and small sample techniques.

7.1 Deep neural networks

Baylon et al. present a multiscale retrosynthesis prediction framework that utilizes a Deep Highway Network (DHN) [107]. This framework operates in two stages: first, a DHN model predicts the reaction group; second, transformation rules for generating the molecule are predicted using DHNs trained on subsets of reactions within the identified reaction group. Hasic et al. developed retrosynthetic models that identify potential breakpoints on molecular substructure fingerprint representations [108]. The model focuses solely on the molecular substructures of the target molecule to pinpoint potential disconnection sites, without relying on additional information such as chemical reaction class. Finally, an essential component of retrosynthetic modeling is a comprehensive pathway evaluation mechanism. To address this, Mo et al. introduced a dynamic tree-structured long short-term memory (tree-LSTM) model [109]. This model enhances the accuracy and reliability of pathway predictions.

7.2 Sequence-to-sequence

The main concept of seq2seq is to frame retrosynthesis prediction as a sequence modeling problem. In this framework, the target molecule serves as the input sequence, while the reactants, reagents, and catalysts comprise the output sequence. Among these models, the Transformer - a seq2seq model based entirely on a multi-head attention mechanism – has emerged as a groundbreaking approach in recent years. Moreover, the development of Bidirectional Encoder Representations from Transformers (BERT) [110] has further enhanced the performance of template-free strategies. Sequence-modeling-based retrosynthesis models, primarily dependent on attention mechanisms, have become the predominant AI models for retrosynthesis.

The integration of chemistry with natural language processing (NLP) was initially proposed by Cadeddu et al [111]. Subsequently, Liu et al. introduced an encoder-decoder framework that utilizes two recurrent neural networks, conceptualizing retrosynthesis prediction as a sequence-to-sequence mapping problem [112]. Compared to template-based baseline models, the Seq2seq model offers several advantages: it can implicitly learn reaction rules and candidate ranking metrics, eliminating the need for distinct reaction complexity ranking metrics used in template-based approaches. Additionally, the seq-2-seq model is more scalable than the rule-based approach. Tetko et al. further advanced the filed by proposing a Transformer model tailored specifically for retrosynthetic reaction prediction tasks [113].

In recent years, Guo et al. introduced a Bayesian inference framework [114] that includes a pretrained Molecular Transformer for forward predictions and employs Bayes' law of conditional probability to reverse the forward model into a backward one. This approach allows for generating a diverse set of highly probable reaction sequences using a Monte Carlo search algorithm combined with backward model. Zheng et al. developed a template-free self-correcting retrosynthesis predictor (SCROP) trained using the Transformer model neural network framework [115]. This model demonstrates superior accuracy for compounds outside its training set compared to other state-of-the-art methods. Duan et al. proposed the attention-based NMT model [116], Tensor2Tensor (T2T), which significantly improves parallelism while reducing training time. Tetko et al. further enhanced the Transformer model framework by employing data augmentation techniques for both input and target data [117], improving prediction capabilities by minimizing the effect of memorized data. Seo et al. proposed Graph Truncated Attention (GTA) [118], which integrates sequence and graph representations by embedding graph information into a seq2seq model. This model adjusts the self-attention layer in the encoder based on the adjacency matrix of the product and modifies the cross-attention layer in the decoder with atomic mappings derived from an automated algorithm. Mann et al. developed a single-step retrosynthetic method using representations based on SMILES grammars [119], with an information-theoretic analysis showing superior performance over traditional SMILES for machine learning tasks. Ucak et al. introduced RetroTRAE, a single-step retrosynthetic prediction method [120] that eschews SMILES-based translation by utilizing fragment and topological descriptors as natural inputs. Wan et al. presented Retroformer [121], a novel structure based on Transformer that avoids conventional cheminformatics tools for molecular editing and employs localized attention to jointly encode molecular sequences and maps. Fang et al. developed a substructure-level decoding model that extracts normally conserved portions of product molecules automatically, employing a fully data-driven approach [122]. Schwaller et al. combined molecular Transformer modeling with hyper-graph exploration strategies to predict reactants, reagents [88], solvents, and catalysts for each retrosynthesis step, and further developed an unsupervised, attention-based network of Transformer models to learn atom mappings [123], bridging the gap between rule-based and data-driven approaches while enhancing chemical interpretability in the prediction results.

The string representation of molecules, particularly using SMILES, faces several challenges, including the generation of invalid strings and insufficient characterization of chemical reactions. To address these issues, Ucak et al. have introduced a novel approach that utilizes molecular fragments combining with template-free sequence-to-sequence models [124], offering a more robust representation of chemical reactions. Zhang et al. have enhanced the molecular transformer models by integrating data expansion and normalized preprocessing strategies [125], which significantly increase accuracy in forward reaction prediction and single-step retrosynthesis across diverse reaction categories. Additionally, Zhong et al. introduced Root-Aligned SMILES (R-SMILES) [126], a method that ensures precise one-to-one mappings between product and reactant SMILES, thereby facilitating more efficient synthesis predictions.

To enhance the diversity of retrosynthesis prediction, Chen et al. introduced a model that broadens the scope of generalizable predictions across various retrosynthetic reactions [127]. This model incorporates two innovative pre-training methods within the Transformer framework and integrates a discrete latent variable model to promote diversity in predictions. Toniato et al. further explained the Transformer’s capabilities by appending a classification token to the language representation of the target molecule [128], which enhances the diversity of the predictions. Meanwhile, Kim et al. developed a dual-path Transformer model that employs a cycle consistency check [129], parameter sharing, and multinomial latent variables, significantly improving accuracy, reducing syntactic errors, and increasing prediction diversity. Irwin et al. introduced Chemformer [130], a Transformer-based model that benefits from self-supervised pre-training, which not only enhances performance but also accelerates convergence in downstream tasks. To address the challenges of making accurate predictions with small chemical datasets, Bai et al. implemented transfer learning techniques in retrosynthesis analysis [131], effectively combining them with seq2seq or Transformer models to improve prediction and validation.

Additionally, the recommendation of reaction conditions remains a critical aspect of retrosynthesis prediction. Addressing this, Andronov et al. proposed a molecular Transformer framework specifically designed to recommend reaction conditions effectively [132].

7.3 Reinforcement learning

Schreck et al. have implemented deep reinforcement learning to optimize reaction path searches, focusing on identifying the most effective reaction strategies during each step of retrosynthesis based on user-defined cost metrics [133]. Their approach involves training a neural network to predict expected synthetic cost using simulation data. Wang et al. introduced a new variant of the Monte Carlo Tree Search (MCTS) variant that enhances the balance between exploration and exploitation within the synthesis space. This model is augmented by a value network, trained through reinforcement learning, and a solvent prediction neural network, which together prove effective in identifying shorter and more environmentally friendly synthetic routes under comparable search conditions.

Building on these advancements, Guo et al. proposed Retrosynthesis Zero (ReSynZ) [134], a method that integrates Monte Carlo Tree Search with reinforcement learning, drawing inspiration from AlphaGo Zero. A key innovation of ReSynZ is its use of complete synthesis paths for complex molecules, derived from reaction rules, as input for neural network training. This allows ReSynZ's neural networks to generate multiple synthesis pathways for a target molecule and suggest potential reaction conditions, even when trained with relatively small datasets. ReSynZ demonstrates strong predictive performance and offers self-improving features, flexible reward settings, and the potential to overcome human limitations in chemical synthesis route planning.

7.4 Graph neural networks

Graph neural networks (GNNs) are specialized deep learning models that excel in processing graph-structured data, particularly for domains such as molecular chemistry, protein interactions, and social networks. In the context of chemistry, GNNs are particularly adept at handling undirected graphs, where atoms are represented as nodes and chemical bonds as edges. This representation aligns well with the inherent structure of chemical molecules, allowing GNNs to capture the complex interactions within molecular structures effectively.

Mao et al. introduced the Graph-enhanced Transformer model (GET), which combines molecular sequence and graph information [135], achieving significantly higher test accuracy than the standard Transformer model. This framework features four different GET designs that integrate SMILES representations with atomic embeddings enhanced by advanced GNNs. Sun et al. proposed a framework that merges sequence-based and graph-based approaches within energy-based models (EBMs), employing different energy functions [136] to highlight connections and differences between models. Additionally, they introduced a new framework that promotes consistency between forward and backward predictions using dual variables. Tu et al. developed Graph2SMILES, a model that leverages the text generation capability of the Transformer model with the permutation invariance of a molecular map encoder, thus minimizing the need for input data augmentation [137]. Liu et al. introduced RetroGNN [138], a novel method for estimating synthesizability by training a GNN with data derived from synthesis planning used on many random molecules, predicting outcomes for a given target molecule. Sacha et al. proposed the Molecular Editing Diagram Attention Network (MEGAN) [139], an end-to-end encoder-decoder model that represents reactions as a series of edits, allowing for effective exploration of plausible chemical reaction spaces. Thakkar et al. introduced prompts describing molecular disconnection to overcome biases in the training database [140], enabling chemists to control disconnection predictions for more diverse and creative retrosynthesis recommendations. Wang et al. introduced RetroExplainer [141], a model that conceptualizes the retrosynthesis task as a molecular assembly process. This model incorporates a multi-meaning and multi-scale graph Transformer, structure-aware contrast learning, and dynamic adaptive multi-task learning, significantly outperforming state-of-the-art methods in single-step inverse synthesis with enhanced interpretability. GNN-Retro [142], developed by Han et al., combines GNNs with advanced search algorithms, utilizing the structure of GNN to incorporate neighboring molecules information for improved estimation accuracy. Jiang et al. enhanced model accuracy by implementing atomic conservation rules through a molecular reconstruction pretraining task and specifying reaction centers through a reaction type-guided comparison pretraining task [143]. Finally, Liu et al. proposed a framework that uses contextual information to improve retrosynthetic planning [144] by viewing synthetic routes as reaction graphs and integrating context through molecule encoding, route information aggregating, and reactant prediction.

7.5 Hybrid AI systems

Recent developments in chemistry-informed search methods have effectively combined modern search algorithms with symbolic AI. Segler et al. introduced the 3N-MCTS model, which incorporates MCTS with an expansion policy network to guide the search [145], and an “in-scope” filter network for pre-selecting the most promising retrosynthetic steps. This method is approximately 30 times faster than traditional search methods relying on extraction rules and hand-coded heuristics, achieving both speed and accuracy. Lin et al. proposed AutoSynRoute, a template-free retrosynthetic model [146] that utilizes a Transformer model for retrosynthesis prediction and MCTS with heuristic scoring for route planning. This model, unlike traditional template-based models, learns the global chemical environments of molecules but shares the inherent limitations of SMILES-based models. Additionally, Hong et al. developed an experience-guided Monte Carlo tree search (EG-MCTS), which leverages synthesis experience rather than standard rollout [147] procedures to optimize the search process. Latendresse et al.’s SynRoute [148] utilizes a relatively small number of reaction templates and a literature-based reaction database to find practical synthetic routes to target compounds, with each reaction template supported by a machine learning classifier for enhanced prediction capabilities. Chen et al. introduced an A* search method using neural network-based models that represent reaction information as AND-OR trees (AND nodes for reactions, OR nodes for molecules), with the search guided by a neural network trained on past retrosynthesis planning experiences to estimate synthesis costs effectively [149]. Chematica [150,151] stands out by utilizing a high-quality chemical database of just 50,000 rules, incorporating strategies to penalize nonselective reactions, strained intermediates, and unlikely structural motifs, along with heuristic searches to efficiently navigate the reaction network. This system terminates its routines upon identifying commercially available building blocks, significantly reducing time and costs by minimizing purification steps. It also includes a bond preservation rule to develop routines distinct from patented alternatives, and notably, Chematica has achieved success in passing the Turing test.

Incorporating robust ranking systems into AI-driven retrosynthesis models can significantly enhance their performance. Lin et al. developed an energy-based model specifically designed to reorder products recommended [152] by existing models such as RetroSim, a similarity-based approach, and NeuralSym, a deep learning approach. This reordering capability notably improves the overall performance of these models. Furthermore, Li et al. introduced RetroRanker [153], a graph neural network-based ranking model that mitigates frequency bias in retrosynthesis predictions by reordering outputs. RetroRanker evaluates potential reaction changes for each predicted reactant set relative to the given products, effectively deprioritizing chemically implausible predictions. Additionally, Jeong et al. proposed ASICS (Advanced System for Intelligent Chemical Synthesis) [154], which utilizes pseudo-A* search strategies to identify optimal synthetic pathways. ASICS minimizes a composite score of synthetic reaction value function, which includes the synthetic accessibility score, likelihood score, and similarity score. This system uniquely balances the exploration between confirmed reaction spaces and unexplored reaction spaces, thereby optimizing the search process for feasible synthetic routes.

Table 3. Overview of retrosynthesis prediction performance for template-free methods.

Methods	**Algorithm**	**Dataset**	**Features**	**TOP-1**	**TOP-5**	**TOP-1**	**TOP-5**	**source code availability**
				with reaction class		**without reaction class**		
Karpov Transformer	transformer	USPTO-50k	SMILES	-	-	42.7	69.8	Y
AutoSynRoute	Transformer +MCTS	USPTO-50k	SMILES	54.6	80.2	43.1	71.8	Y
Bayesian-Retro (MT-predictable)	Transformer +SMC	USPTO-50k	SMILES	62.1	88.8	53.8	84.1	N
Chemformer	transformer	USPTO-50k	SMILES	-	-	54.3	62.3	Y
tree-LSTM	LSTM	Pistachio+ ASKCOS	fingerprint	-	-	79.1	88.6	N
G2Retro	MPN	USPTO-50k	graph	63.6	88.4	54.1	81.2	Y
GTA	attention mechanism	USPTO-50k	SMILES	-	-	51.1	74.8	N
LV-transformer	transformer	USPTO-50k	SMILES	-	-	40.5	72.8	N
GTE	GNN+trans former	USPTO-full	graph	76.6	89.6	44.9	62.4	Y
MEGAN	GAN	USPTO-50k	graph	60.7	87.5	48.1	78.4	Y
Graph2SMILES	GNN	USPTO-full	graph	-	-	52.9	70	Y
SCROP	transformer	USPTO-50k	SIMLES	59	78.1	43.7	65.2	N
Retroformer	transformer	USPTO-50k	SMILES	64	86.7	53.2	76.6	Y
SMILES-grammar-based	transformer	USPTO-50k	SIMLES-like	43.8	61.4	32.1	48.9	N
T2T	attention mechanism	USPTO-50k	SMILES	-	-	51	69	Y
RetroTRAE	transformer	USPTO-full	ECFP	-	-	58.3	-	Y
Liu	seq2seq	USPTO-50k	SMILES	-	-	37.4	57	Y
Molecular Substructure	-	USPTO-50k	HSFP	61.4	70.4	61.4	70.4	Y
Fang	transformer	USPTO-full	SMILES	-	-	50.4	-	Y
AT	transformer	USPTO-50k	SMILES	-	-	53.5	81	Y
Substructure-based	seq2seq	USPTO-full	MACCS	-	-	29	-	Y
Dual-TF	GNN+seq2seq	USPTO-50k	graph+ SMILES	65.7	84.7	53.6	74.6	N
seq2seq-transfer learning	transfer learning	USPTO-50k	SMILES	-	-	60.7	83.5	N
Two-way transformers	transformer	USPTO-50k	SMILES	-	-	47.1	73.1	Y
RetroExplainer	Graph Transformer	USPTO-50k	graph	66.8	92.5	57.7	84.8	Y
R-SMILES	transformer	USPTO-50k	SMILES	-	-	56.3	86.2	Y
Zhang et al.	transformer	USPTO-50k	SMILES	55	79	43	73	N
Pre-training transformer	transformer	USPTO-50k	SMILES	67.1	85.2	62	78.4	N

8. Semi-template-based models

Semi-template-based methods offer an alternative approach to retrosynthesis that does not rely directly on reaction templates for transforming products into its reactants. Instead, these methods employ a two-step workflow utilizing atom-mappings: first identifying reaction centers and transforming the product into synthons (intermediate molecules); and second, completing these synthons into reactants.

Shi et al. introduced G2Gs [79], a method that segments the target molecular map into synthons by identifying reaction centers and translates these synthons into the final reactant maps through a variogram translation framework. This approach outperforms traditional template-based methods like RetroSim [96] and Neuralsym [97]. Similarly, Chen et al. developed the one-step retrosynthesis prediction framework G2Retro, which predicts reactive centers in a target molecule, identifies synthons to assemble the target, and converts these synthons into reactants, thereby defining a comprehensive set of reactive center types. Nicolaou et al. introduced a chemical context-aware data-driven method based on DDRAM algorithm, which recommends synthetic routes matching a precedent-derived template [155]. Yan et al. proposed RetroXpert [156], which decomposes retrosynthesis into two steps: identifying potential reaction centers in the target molecule using graph neural networks, and predicting relevant reactants based on the obtained synthetics with a reactant generation model. Wang et al. developed RetroPrime, a single-step, template-free approach based on the Transformer model. [157]. This method decomposes a molecule into a synthon and then generates a reactant by attaching leaving groups, utilizing a generalized Transformer model. Somnath et al. propose a graph-based approach that assumes the graph topology of precursor molecules remains invariant during chemical reactions [158]. The model first predicts a set of graph edits that transform the target into a synthon, which are then expanded into molecules. ReTReK, introduced by Ishida et al., is a data-driven and rule-based retrosynthesis model that formlates four scores for synthesis route evaluation, using a GCN and MCTS for retrosynthesis prediction and path search, respectively [159]. Zhang et al. employ a chemistry-informed molecular graph (CIMG) as a molecular representation [160], which includes features like NMR chemical shifts, bond dissociation energies, and solvent/catalyst information, and uses Message Passing Neural Networks (MPNN)-layered GNN models to select reaction templates, infer reactants, and verify the plausibility of the proposed reaction. Lin et al. proposed a graph-to-graph transformation model, G2GT [161], which is built on the standard Transformer model structure with data augmentation and enhanced by a weak ensemble approach that combines beam search, kernel, and top-k sampling methods to boost diversity. Finally, Zhong et al. propose Graph2Edits, an end-to-end framework [162] based on a graphical neural network that predicts edits of a product graph in an auto-regressive manner, thereby streamlining the transformation of intermediates and final reactants in a unified learning process.

Table 4. Overview of retrosynthesis prediction performance for semi-template-based methods.

Methods	Algorithm	dataset	features	TOP-1	TOP-5	TOP-1	TOP-5	source code availability
				with reaction class		without reaction class		
G2Gs	GCN	USPTO-50k	graph	61	86	48.9	72.5	N
ReTReK	GCN+MCTS	Reaxys	SMILES	-	-	36.1	-	Y
G2GT	GNN+transformer	USPTO-50k	graph	-	-	54.1	74.5	N
GraphRetro	MPN	USPTO-50k	graph	63.9	85.2	53.7	72.2	Y
RetroPrime	transformer	USPTO-50k	SMILES	64.8	81.6	51.4	74	Y
Graph2Edits	GNN	USPTO-50k	graph	67.1	91.5	55.1	83.4	Y
RetroXpert	GNN	USPTO-50k	graph	62.1	75.8	50.4	62.3	N

9. Comparison of three categorizations

Top-k accuracy is a vital metric for evaluating single-step retrosynthesis models, especially given the multiplicity of viable pathways in organic synthesis. Therefore, it is important to assess model performance using both top-1 and top-5 accuracy metrics to avoid misleading conclusions that might arise from focusing solely on top-1 accuracy.

As shown in **Figure 9**, template-based and semi-template-based models demonstrate higher average accuracy in scenarios where the reaction class is known. These models also maintain relatively high average accuracy even when the reaction class is unspecified, highlighting their robustness. Further analysis of data distribution reveals that both template-based and semi-template-based models feature more tightly clustered distributions, suggesting greater stability in their predictive performance. In contrast, template-free models show higher dispersion and lower stability, which underscores the significant variability in their performance.

In conclusion, while template-based models demonstrate consistently high accuracy and stability, semi-template-based models also show promising results and the potential to lead in performance. However, the substantial variability observed in template-free methods calls for cautious application and ongoing development to overcome the challenges in AI-assisted retrosynthesis.

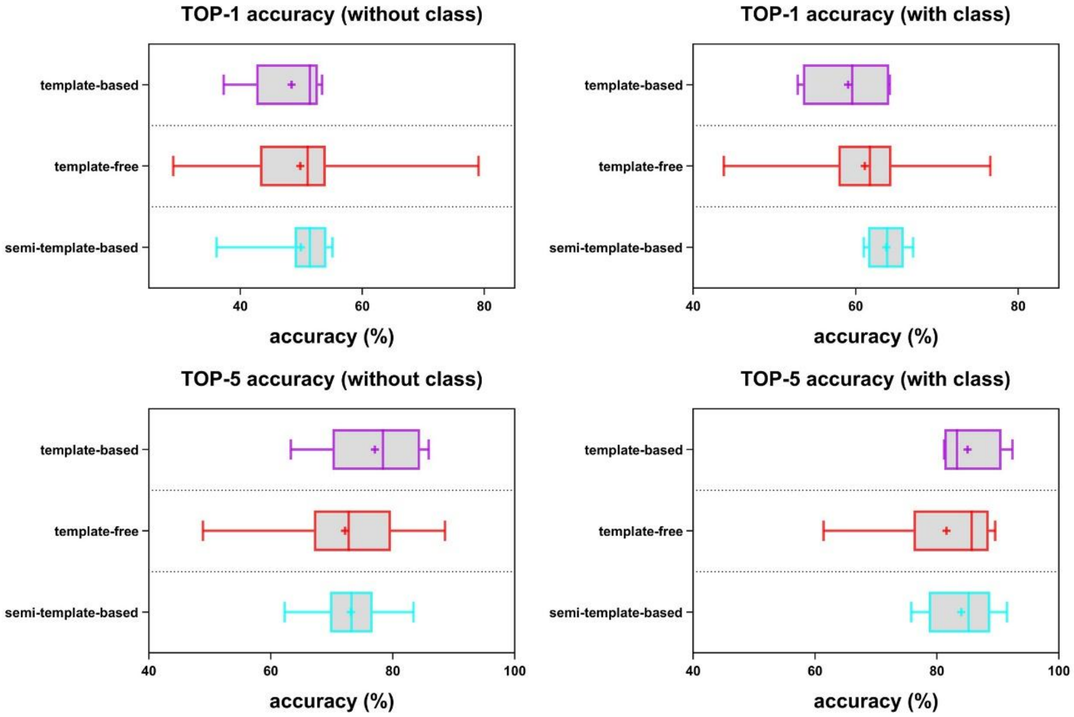

Figure 9. TOP-k accuracy for template-based, template-free, and semi-template-based methods.

10. Challenges and future directions for retrosynthesis prediction researchers

In recent years, the rapid development of AI-assisted retrosynthesis models has been remarkable, yet there remain significant challenges and limitations that need addressing:

1. Data Quality and Diversity: One of the foremost challenges is the lack of sufficient high-quality data. The development of high-performance AI models is contingent upon the availability of both substantial quantity and quality of training data. Public datasets are often limited compared to commercial databases. There is a critical need for collaboration among computational chemists to prepare expansive and diverse datasets. These datasets should include varied information such as stereochemistry information, solvents, and catalysts. The creation of a comprehensive chemical reaction database requires developing standardized methods for managing and integrating diverse data sources.

2. Model Interpretability: Purely data-driven models frequently suffer from a lack of interpretability, making it difficult for researchers to understand the underlying reasons behind predictions. Employing model-agnostic techniques like LIME [163–165], SHAP [163], and Anchors [166] can enhance interpretability by providing both global and local explanations while highlighting key features influence the model decisions. Furthermore, the adoption of explainable neural networks [167,168], which incorporate interpretable layers via attention and gating mechanisms, is strongly encouraged.

3. Template-free Model Limitations: Template-free models, often trained on text sequences, may overlook the chemical significance of bond disconnection, which can occasionally result in impractical suggestions. These models may also exhibit bias due stemming from the underrepresentation of rare reactions in training datasets. A promising approach to mitigate these issues is to blend data-driven methods with fundamental chemical principles to reduce bias and improve the model's utility.

4. Experimental Validation: For any in silico design process, suggested synthesis routes must be validated experimentally. High-throughput and parallelized experimentation are vital for rapid data generation and validation. However, the lack of experimental conditions in most retrosynthesis prediction models adds an extra layer of challenge to experimental planning. Recent advancements in automated experimental design, which leverage AI to optimize and pinpoint feasible reaction conditions [169,170], are crucial steps forward.

Given these challenges, several promising research directions emerge for further retrosynthesis prediction:

1. Developing high-quality and diverse chemical reaction datasets and creating intelligent algorithms capable of handling incomplete or inaccurate data represent fundamental research priorities.

2. Focusing on model interpretability and visualization will likely be a key research direction. Integrating design of experiments (DoE) with robotic experimental steps is also essential.

3. It is recommended to create more intricate and comprehensive models that can encompass a wider range of reaction types and conditions.

4. Exploring the synergy between AI algorithms and traditional rules could yield substantial improvement.

5. Greater emphasis should be placed on developing more efficient and environmentally friendly chemical synthesis conditions within retrosynthesis predictions.

11. Conclusion

CASP studies play an integral role in enhancing drug design, significantly accelerating the synthesis process and reducing associated costs. With the advent of modern intelligent algorithms, there is substantial potential to further enhance the efficiency and accuracy of CASP. Future research in this area should prioritize the development of robust, interpretable retrosynthesis models and the extraction of higher-quality chemical reaction datasets from patents and literature. A pivotal aspect of CASP research is ensuring the interpretability of AI analyses, which is essential for improving the transparency and reliability of AI-driven predictions. The efficacy of data-oriented methods heavily relies on the quality of the reaction databases, thus, ensuring high-quality datasets is indispensable. Moreover, interdisciplinary collaborations among computer scientists, statisticians, organic chemists, and computational chemists are becoming increasingly vital. These collaborations can merge diverse perspectives and expertise to address complex organic retrosynthesis tasks effectively. The state of AI-assisted synthetic planning is still evolving, and comprehensive assessments are necessary to fully evaluate its potential. Due to variations in training datasets, direct comparisons of AI model performances using identical evaluation metrics may not always be feasible, and no single model consistently outperforms others across all tasks.

This review offers an extensive overview of the latest developments in CASP driven by intelligent algorithms, categorizing existing models into three main types: template-based, template-free, and semi-template-based models. Our comparative analysis reveals that semi-template-based models typically exhibit superior performance. We also outline the major challenges currently facing this field and suggest future directions for CASP research. Recent studies highlight the significant potential of artificial intelligence in retrosynthetic prediction, which could alleviate the time and cost burdens on organic chemists in synthesis planning. By the end of this review, it is our hope that scientists engaged in AI-assisted retrosynthesis prediction will find the insights needed to select methodologies that best align with their research objectives. The discussions provided here should inspire future enhancements and explorations in the field. As retrosynthesis techniques continue to mature, we anticipate their integration into automated chemical synthesis systems [171], which could revolutionize the manufacturing of chemical compounds and have profound social and technological impacts.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

This work was supported by the National Key R&D Program of China (Grant No. 2023YFF1204903), the National Natural Science Foundation of China (Grant Nos. 22222303, 22173032, 22361132538 and 92461313), and the Artificial Intelligence-Driven Reform of Scientific Research Paradigms: Empowerment Program for Discipline Advancement (Grant No. 2024AI01009). We also thank the ECNU Multifunctional Platform for Innovation (No. 001) for providing supercomputer time. T.Z. dedicates this work to Professor Keli Han, whose unwavering support, guidance, and kindness over the past decade have profoundly shaped his career and character. May his generous spirit and enduring mentorship continue to inspire us, and may he rest in peace.

References

[1] Nicolaou K.C. and Chen J.S., The art of total synthesis through cascade reactions. *Chem. Soc. Rev.*, **38** (2009), 2993.

[2] Nantermet P.G., Reaction: the art of synthetic chemistry. *Chem.*, **1** (2016), 335–336.

[3] Gordon E.M., Gallop M.A. and Patel D.V., Strategy and tactics in combinatorial organic synthesis. *Applications. to. Drug. Discovery. Acc. Chem. Res.*, **29** (1996), 144–154.

[4] Yeh B.J. and Lim W.A., Synthetic biology: lessons from the history of synthetic organic chemistry. *Nat. Chem. Biol.*, **3** (2007), 521–525.

[5] Campos K.R., Coleman P.J., Alvarez J.C., Dreher S.D., Garbaccio R.M. and Terrett N.K., et al., The importance of synthetic chemistry in the pharmaceutical industry. *Science*, **363** (2019), eaat0805.

[6] Blakemore D.C., Castro L., Churcher I., Rees D.C., Thomas A.W. and Wilson D.M., et al., Organic synthesis provides opportunities to transform drug discovery. *Nat. Chem.*, **10** (2018), 383–394.

[7] Corey E.J., Robert robinson lecture. *Retrosynthetic. thinking—essentials. and. examples. Chem,. Soc,. Rev.*, **17** (1988), 111–133.

[8] Bergman R.G. and Danheiser R.L., Reproducibility in chemical research. *Angew. Chem. Int. Ed.*, **55** (2016), 12548–12549.

[9] Duros V., Grizou J., Xuan W., Hosni Z., Long D.-L. and Miras H.N., et al., Human versus robots in the discovery and crystallization of gigantic polyoxometalates. *Angew. Chem. Int. Ed.*, **56** (2017), 10815–10820.

[10] Ley S.V., The engineering of chemical synthesis: humans and machines working in harmony. *Angew. Chem. Int. Ed.*, **57** (2018), 5182–5183.

[11] PENSAK D.A. and COREY E.J., Lhasa—logic and heuristics applied to synthetic analysis. *Computer. Assisted. Organic. Synthesis,. vol. 61,. AMERICAN CHEMICAL SOCIETY*, **1** (1977).

[12] Henson A.B., Gromski P.S. and Cronin L., Designing algorithms to aid discovery by chemical robots. *ACS Cent. Sci.*, **4** (2018), 793–804.

[13] Lusher S.J., McGuire R., Van Schaik R.C., Nicholson C.D. and De Vlieg J., Data-driven medicinal chemistry in the era of big data. *Drug. Discovery. Today.*, **19** (2014), 859–868.

[14] Tetko I.V., Engkvist O., Koch U., Reymond J. and Chen H., Bigchem: challenges and opportunities for big data analysis in chemistry. *Mol. Inf.*, **35** (2016), 615–621.

[15] Hessler G. and Baringhaus K.-H., Artificial intelligence in drug design. *Molecules.*, **23** (2018), 2520.

[16] Sellwood M.A., Ahmed M., Segler M.H. and Brown N., Artificial intelligence in drug discovery. *Future. Med. Chem.*, **10** (2018), 2025–2028.

[17] Cho K., van Merrienboer B., Gulcehre C., Bahdanau D., Bougares F. and Schwenk H., et al., Learning phrase representations using RNN encoder-decoder for statistical machine translation 2014. .

[18] Sutskever I., Vinyals O. and Le Q.V., Sequence to sequence learning with neural networks. *Advances. in. Neural. Information. Processing. Systems,. vol. 27,. Curran. Associates,. Inc.*, (2014).

[19] Bahdanau D., Cho K. and Bengio Y., Neural machine translation by jointly learning to align and translate 2014. .

[20] Gehring J., Auli M., Grangier D., Yarats D. and Dauphin Y.N., Convolutional sequence to sequence learning. *Proceedings. of. the. 34th. International. Conference. on. Machine. Learning,. PMLR*, **1243** (2017).

[21] Gori M., Monfardini G. and Scarselli F., A new model for learning in graph domains. *Proceedings.*, **2005** (2005), vol.

[22] Sperduti A. and Starita A., Supervised neural networks for the classification of structures. *IEEE Trans. Neural. Netw.*, **8** (1997), 714–735.

[23] Gallicchio C. and Micheli A., Graph echo state networks. *The.*, **2010** (2010), p.

[24] Scarselli F., Gori M., Tsoi A.C., Hagenbuchner M. and Monfardini G., The graph neural network model. *IEEE Transactions. on. Neural. Networks.*, **20** (2009), 61–80.

[25] Kipf T.N. and Welling M., Semi-supervised classification with graph convolutional networks 2016. .

[26] Veličković P., Cucurull G., Casanova A., Romero A., Liò P. and Bengio Y., Graph attention networks 2017. .

[27] Li Y., Tarlow D., Brockschmidt M. and Zemel R., Gated graph sequence neural networks 2015. .

[28] Cao S., Lu W. and Xu Q., Deep neural networks for learning graph representations. *Proc. AAAI Conf. Artif. Intell.*, **30** (2016).

[29] De Cao N. and Kipf T., Molgan: an implicit generative model for small molecular graphs 2018. .

[30] Yan S., Xiong Y. and Lin D., Spatial temporal graph convolutional networks for skeleton-based action recognition. *Proc. AAAI Conf. Artif. Intell.*, **32** (2018).

[31] Kaelbling L.P., Littman M.L. and Moore A.W., Reinforcement learning: A survey. *J. Artificial. Intelligence. Res.*, **4** (1996), 237–285.

[32] Lillicrap T.P., Hunt J.J., Pritzel A., Heess N., Erez T. and Tassa Y., et al., Continuous control with deep reinforcement learning 2019. .

[33] Mnih V., Badia A.P., Mirza A., Graves A., Lillicrap T. and Harley T., et al., Asynchronous methods for deep reinforcement learning. *Proceedings. of. The. 33rd. International. Conference. on. Machine. Learning,. PMLR*, **1928** (2016).

[34] Mnih V., Kavukcuoglu K., Silver D., Graves A., Antonoglou I. and Wierstra D., et al., Playing atari with deep reinforcement learning 2013. .

[35] Hessel M., Modayil J., Van Hasselt H., Schaul T., Ostrovski G. and Dabney W., et al., Rainbow: combining improvements in deep reinforcement learning. *Proc. AAAI Conf. Artif. Intell.*, **32** (2018).

[36] Moerland T.M., Broekens J., Plaat A. and Jonker C.M., Model-based reinforcement learning: A survey 2022. .

[37] Nair A., Pong V., Dalal M., Bahl S., Lin S. and Levine S., Visual reinforcement learning with imagined goals 2018. .

[38] Kulkarni T.D., Narasimhan K., Saeedi A. and Tenenbaum J., Hierarchical deep reinforcement learning: integrating temporal abstraction and intrinsic motivation. *Advances. in. Neural. Information. Processing. Systems,. vol. 29,. Curran. Associates,. Inc.*, (2016).

[39] Horling B. and Lesser V., A survey of multi-agent organizational paradigms. *Knowl. Eng. Rev.*, **19** (2004), 281–316.

[40] Schrittwieser J., Antonoglou I., Hubert T., Simonyan K., Sifre L. and Schmitt S., et al., Mastering atari, go, chess and shogi by planning with a learned model. *Nature*, **588** (2020), 604–609.

[41] OpenAI, Akkaya I., Andrychowicz M., Chociej M., Litwin M. and McGrew B., et al., Solving rubik's cube with a robot hand 2019. .

[42] Sallab A.E., Abdou M., Perot E. and Yogamani S., Deep reinforcement learning framework for autonomous driving. *Electron. Imaging.*, **29** (2017), 70–76.

[43] Jeon W. and Kim D., Autonomous molecule generation using reinforcement learning and docking to develop potential novel inhibitors. *Sci. Rep.*, **10** (2020), 22104.

[44] Silver D., Singh S., Precup D. and Sutton R.S., Reward is enough. *Artif. Intell.*, **299** (2021), 103535.

[45] Speech Understanding S.Y.S.T.E.M.S., Summary of results of the five-year research effort at carnegie-mellon university. (n.d.).

[46] Zeng W. and Church R.L., Finding shortest paths on real road networks: the case for a*. *Int. J. Geogr. Inf. Sci.*, **23** (2009), 531–543.

[47] Coulom R., Efficient selectivity and backup operators in monte-carlo tree search. *In:. Van. Den. Herik. HJ,. Ciancarini. P,. Donkers. HHLM,. editors. Computers. and. Games,. vol. 4630,. Berlin,. Heidelberg:. Springer.*, **72** (2007).

[48] Silver D., Huang A., Maddison C.J., Guez A., Sifre L. and Van Den Driessche G., et al., Mastering the game of go with deep neural networks and tree search. *Nature*, **529** (2016), 484–489.

[49] Rich A.S. and Gureckis T.M., Lessons for artificial intelligence from the study of natural stupidity. *Nat. Mach. Intell.*, **1** (2019), 174–180.

[50] He J., Chemu dataset for information extraction from chemical patents 2020;2. .

[51] Lowe D.M., Extraction of chemical structures and reactions from the literature 2012. .

[52] Schneider N., Stiefl N. and Landrum G.A., What's what: the (nearly) definitive guide to reaction role assignment. *J. Chem. Inf. Model.*, **56** (2016), 2336–2346.

[53] Jin W., Coley C., Barzilay R. and Jaakkola T., Predicting organic reaction outcomes with weisfeiler-lehman network. *Advances. in. Neural. Information. Processing. Systems,. vol. 30,. Curran. Associates,. Inc.*, (2017).

[54] Jia X., Lynch A., Huang Y., Danielson M., Lang'at I. and Milder A., et al., Anthropogenic biases in chemical reaction data hinder exploratory inorganic synthesis. *Nature*, **573** (2019), 251–255.

[55] Coley C.W., Eyke N.S. and Jensen K.F., Autonomous discovery in the chemical sciences part ii: outlook. *Angew. Chem. Int. Ed.*, **59** (2020), 23414–23436.

[56] Lin S., Dikler S., Blincoe W.D., Ferguson R.D., Sheridan R.P. and Peng Z., et al., Mapping the dark space of chemical reactions with extended nanomole synthesis and MALDI-TOF MS. *Science*, **361** (2018), eaar6236.

[57] Vaucher A.C., Zipoli F., Geluykens J., Nair V.H., Schwaller P. and Laino T., Automated extraction of chemical synthesis actions from experimental procedures. *Chemistry.*, (2019).

[58] Drake D., ELN implementation challenges. *Drug. Discovery. Today.*, **12** (2007), 647–649.

[59] Saebi M., Nan B., Herr J.E., Wahlers J., Guo Z. and Zurański A.M., et al., On the use of real-world datasets for reaction yield prediction. *Chem. Sci.*, **14** (2023), 4997–5005.

[60] Thakkar A., Kogej T., Reymond J.-L., Engkvist O. and Bjerrum E.J., Datasets and their influence on the development of computer assisted synthesis planning tools in the pharmaceutical domain. *Chem. Sci.*, **11** (2020), 154–168.

[61] Weininger D., Smiles, a chemical language and information system. *1. Introduction. to. methodology. and. encoding. rules. J. Chem. Inf. Comput. Sci.*, **28** (1988), 31–36.

[62] Weininger D., Weininger A. and Weininger J.L., SMILES. *2. Algorithm. for. generation. of. unique. SMILES notation. J. Chem. Inf. Comput. Sci.*, **29** (1989), 97–101.

[63] Zhou G., Gao Z., Ding Q., Zheng H., Xu H. and Wei Z., et al., Uni-mol: A universal 3D molecular representation learning framework 2023. .

[64] Cadeddu A., Wylie E.K., Jurczak J., Wampler-Doty M. and Grzybowski B.A., Organic chemistry as a language and the implications of chemical linguistics for structural and retrosynthetic analyses. *Angew. Chem. Int. Ed.*, **53** (2014), 8108–8112.

[65] Krenn M., Häse F., Nigam A., Friederich P. and Aspuru-Guzik A., Self-referencing embedded strings (selfies): A 100% robust molecular string representation. *Mach. Learn:. Sci. Technol.*, **1** (2020), 045024.

[66] Heller S.R., McNaught A., Pletnev I., Stein S. and Tchekhovskoi D., Inchi, the IUPAC international chemical identifier. *J. Cheminf.*, **7** (2015), 23.

[67] Cereto-Massagué A., Ojeda M.J., Valls C., Mulero M., Garcia-Vallvé S. and Pujadas G., Molecular fingerprint similarity search in virtual screening. *Methods.*, **71** (2015), 58–63.

[68] Durant J.L., Leland B.A., Henry D.R. and Nourse J.G., Reoptimization of MDL keys for use in drug discovery. *J. Chem. Inf. Comput. Sci.*, **42** (2002), 1273–1280.

[69] Bolton E.E., Wang Y., Thiessen P.A. and Bryant S.H., Chapter 12 - pubchem: integrated platform of small molecules and biological activities. *In:. Wheeler. RA,. Spellmeyer. DC,. editors. Annual. Reports. in. Computational. Chemistry,. vol. 4,. Elsevier.*, **217** (2008).

[70] Barnard J.M. and Downs G.M., Chemical fragment generation and clustering software. *J. Chem. Inf. Comput. Sci.*, **37** (1997), 141–142.

[71] Tovar A., Eckert H. and Bajorath J., Comparison of 2D fingerprint methods for multiple-template similarity searching on compound activity classes of increasing structural diversity. *ChemMedChem.*, **2** (2007), 208–217.

[72] Sheridan R.P., Miller M.D., Underwood D.J. and Kearsley S.K., Chemical similarity using geometric atom pair descriptors. *J. Chem. Inf. Comput. Sci.*, **36** (1996), 128–136.

[73] Daylight Theory: Fingerprints N., D. .

[74] Bender A., Mussa H.Y., Glen R.C. and Reiling S., Similarity searching of chemical databases using atom environment descriptors (molprint 2d): evaluation of performance. *J. Chem. Inf. Comput. Sci.*, **44** (2004), 1708–1718.

[75] Bender A., Mussa H.Y., Glen R.C. and Reiling S., Molecular similarity searching using atom environments, information-based feature selection, and a naïve bayesian classifier. *J. Chem. Inf. Comput. Sci.*, **44** (2004), 170–178.

[76] Rogers D. and Hahn M., Extended-connectivity fingerprints. *J. Chem. Inf. Model.*, **50** (2010), 742–754.

[77] Schwalbe-Koda D. and Gómez-Bombarelli R., Generative models for automatic chemical design. *In:. Schütt. KT,. Chmiela. S,. von. Lilienfeld. OA,. Tkatchenko. A,. Tsuda. K,. Müller. K. R,. editors. Machine. Learning. Meets. Quantum. Physics,. Cham:. Springer. International. Publishing.*, **445** (2020).

[78] Kearnes S., McCloskey K., Berndl M., Pande V. and Riley P., Molecular graph convolutions: moving beyond fingerprints. *J. Comput. Aided. Mol. Des.*, **30** (2016), 595–608.

[79] Shi C., Xu M., Guo H., Zhang M. and Tang J., A graph to graphs framework for retrosynthesis prediction. *Proceedings. of. the. 37th. International. Conference. on. Machine. Learning,. PMLR*, **8818** (2020).

[80] Kwon Y., Lee D., Choi Y.-S., Shin K. and Kang S., Compressed graph representation for scalable molecular graph generation. *J. Cheminf.*, **12** (2020), 58.

[81] Varnek A., Fourches D., Hoonakker F. and Solov'ev V.P., Substructural fragments: an universal language to encode reactions, molecular and supramolecular structures. *J. Comput. Aided. Mol. Des.*, **19** (2005), 693–703.

[82] Fukunishi Y., Kurosawa T., Mikami Y. and Nakamura H., Prediction of synthetic accessibility based on commercially available compound databases. *J. Chem. Inf. Model.*, **54** (2014), 3259–3267.

[83] Ertl P. and Schuffenhauer A., Estimation of synthetic accessibility score of drug-like molecules based on molecular complexity and fragment contributions. *J. Cheminf.*, **1** (2009), 8.

[84] Molga K., Szymkuć S. and Grzybowski B.A., Chemist ex machina: advanced synthesis planning by computers. *Acc. Chem. Res.*, **54** (2021), 1094–1106.

[85] Coley C.W., Rogers L., Green W.H. and Jensen K.F., Scscore: synthetic complexity learned from a reaction corpus. *J. Chem. Inf. Model.*, **58** (2018), 252–261.

[86] Podolyan Y., Walters M.A. and Karypis G., Assessing synthetic accessibility of chemical compounds using machine learning methods. *J. Chem. Inf. Model.*, **50** (2010), 979–991.

[87] Li J. and Eastgate M.D., Current complexity: a tool for assessing the complexity of organic molecules. *Org. Biomol. Chem.*, **13** (2015), 7164–7176.

[88] Schwaller P., Petraglia R., Zullo V., Nair V.H., Haeuselmann R.A. and Pisoni R., et al., Predicting retrosynthetic pathways using transformer-based models and a hyper-graph exploration strategy. *Chem. Sci.*, **11** (2020), 3316–3325.

[89] Bender A. and Glen R.C., Molecular similarity: A key technique in molecular informatics. *Org. Biomol. Chem.*, **2** (2004), 3204.

[90] Szymkuć S., Gajewska E.P., Klucznik T., Molga K., Dittwald P. and Startek M., et al., Computer-assisted synthetic planning: the end of the beginning. *Angew. Chem. Int. Ed.*, **55** (2016), 5904–5937.

[91] Llanos E.J., Leal W., Luu D.H., Jost J., Stadler P.F. and Restrepo G., Exploration of the chemical space and its three historical regimes. *Proc. Natl. Acad. Sci.*, **116** (2019), 12660–12665.

[92] Coley C.W., Green W.H. and Jensen K.F., Rdchiral: an rdkit wrapper for handling stereochemistry in retrosynthetic

template extraction and application. *J. Chem. Inf. Model.*, **59** (2019), 2529–2537.

[93] Jaworski W., Szymkuć S., Mikulak-Klucznik B., Piecuch K., Klucznik T. and Kaźmierowski M., et al., Automatic mapping of atoms across both simple and complex chemical reactions. *Nat. Commun.*, **10** (2019), 1434.

[94] Plehiers P.P., Marin G.B., Stevens C.V. and Van Geem K.M., Automated reaction database and reaction network analysis: extraction of reaction templates using cheminformatics. *J. Cheminf.*, **10** (2018), 11.

[95] Plehiers P.P., Marin G.B., Stevens C.V. and Van Geem K.M., Automated reaction database and reaction network analysis: extraction of reaction templates using cheminformatics. *J. Cheminf.*, **10** (2018), 11.

[96] Coley C.W., Rogers L., Green W.H. and Jensen K.F., Computer-assisted retrosynthesis based on molecular similarity. *ACS Cent. Sci.*, **3** (2017), 1237–1245.

[97] Segler M.H.S. and Waller M.P., Neural-symbolic machine learning for retrosynthesis and reaction prediction. *Chemistry. A. European. J.*, **23** (2017), 5966–5971.

[98] Watson I.A., Wang J. and Nicolaou C.A., A retrosynthetic analysis algorithm implementation. *J. Cheminf.*, **11** (2019), 1.

[99] Genheden S., Thakkar A., Chadimová V., Reymond J.-L., Engkvist O. and Bjerrum E., Aizynthfinder: a fast, robust and flexible open-source software for retrosynthetic planning. *J. Cheminf.*, **12** (2020), 70.

[100] Park M.S., Lee D., Kwon Y., Kim E. and Choi Y.-S., Data undersampling models for the efficient rule-based retrosynthetic planning. *Phys. Chem. Chem. Phys.*, **23** (2021), 26510–26518.

[101] Chen S. and Jung Y., Deep retrosynthetic reaction prediction using local reactivity and global attention. *JACS Au.*, **1** (2021), 1612–1620.

[102] Seidl P., Renz P., Dyubankova N., Neves P., Verhoeven J. and Wegner J.K., et al., Improving few- and zero-shot reaction template prediction using modern hopfield networks. *J. Chem. Inf. Model.*, **62** (2022), 2111–2120.

[103] Genheden S., Norrby P.-O. and Engkvist O., Aizynthtrain: robust, reproducible, and extensible pipelines for training synthesis prediction models. *J. Chem. Inf. Model.*, **63** (2023), 1841–1846.

[104] Dai H., Li C., Coley C.W., Dai B. and Song L., Retrosynthesis prediction with conditional graph logic network 2020. .

[105] Yan C., Zhao P., Lu C., Yu Y. and Huang J., Retrocomposer: composing templates for template-based retrosynthesis prediction. *Biomolecules.*, **12** (2022), 1325.

[106] Toniato A., Unsleber J.P., Vaucher A.C., Weymuth T., Probst D. and Laino T., et al., Quantum chemical data generation as fill-in for reliability enhancement of machine-learning reaction and retrosynthesis planning. *Digit. Discov.*, **2** (2023), 663–673.

[107] Baylon J.L., Cilfone N.A., Gulcher J.R. and Chittenden T.W., Enhancing retrosynthetic reaction prediction with deep learning using multiscale reaction classification. *J. Chem. Inf. Model.*, **59** (2019), 673–688.

[108] Hasic H. and Ishida T., Single-step retrosynthesis prediction based on the identification of potential disconnection sites using molecular substructure fingerprints. *J. Chem. Inf. Model.*, **61** (2021), 641–652.

[109] Mo Y., Guan Y., Verma P., Guo J., Fortunato M.E. and Lu Z., et al., Evaluating and clustering retrosynthesis pathways with learned strategy. *Chem. Sci.*, **12** (2021), 1469–1478.

[110] Schwaller P., Probst D., Vaucher A.C., Nair V.H., Kreutter D. and Laino T., et al., Mapping the space of chemical reactions using attention-based neural networks. *Nat. Mach. Intell.*, **3** (2021), 144–152.

[111] Cadeddu A., Wylie E.K., Jurczak J., Wampler-Doty M. and Grzybowski B.A., Organic chemistry as a language and the implications of chemical linguistics for structural and retrosynthetic analyses. *Angew. Chem. Int. Ed.*, **53** (2014), 8108–8112.

[112] Liu B., Ramsundar B., Kawthekar P., Shi J., Gomes J. and Luu Nguyen Q., et al., Retrosynthetic reaction prediction using neural sequence-to-sequence models. *ACS Cent. Sci.*, **3** (2017), 1103–1113.

[113] Karpov P., Godin G. and Tetko I.V., A transformer model for retrosynthesis. *In:. Tetko. IV,. Kůrková. V,. Karpov. P,. Theis. F,. editors. Artificial. Neural. Networks. and. Machine. Learning. – ICANN*, **11731** (2019), Springer International Publishing; 2019.

[114] Guo Z., Wu S., Ohno M. and Yoshida R., Bayesian algorithm for retrosynthesis. *J. Chem. Inf. Model.*, **60** (2020), 4474–4486.

[115] Zheng S., Rao J., Zhang Z., Xu J. and Yang Y., Predicting retrosynthetic reactions using self-corrected transformer neural networks. *J. Chem. Inf. Model.*, **60** (2020), 47–55.

[116] Duan H., Wang L., Zhang C., Guo L. and Li J., Retrosynthesis with attention-based NMT model and chemical analysis of “wrong” predictions. *RSC Adv.*, **10** (2020), 1371–1378.

[117] Tetko I.V., Karpov P., Van Deursen R. and Godin G., State-of-the-art augmented NLP transformer models for direct and single-step retrosynthesis. *Nat. Commun.*, **11** (2020), 5575.

[118] Seo S.-W., Song Y.Y., Yang J.Y., Bae S., Lee H. and Shin J., et al., Gta: graph truncated attention for retrosynthesis. *AAAI*, **35** (2021), 531–539.

[119] Mann V. and Venkatasubramanian V., Retrosynthesis prediction using grammar-based neural machine translation: an information-theoretic approach. *Comput. Chem. Eng.*, **155** (2021), 107533.

[120] Ucak U.V., Ashyrmamatov I., Ko J. and Lee J., Retrosynthetic reaction pathway prediction through neural machine translation of atomic environments. *Nat. Commun.*, **13** (2022), 1186.

[121] Wan Y., Hsieh C.-Y., Liao B. and Zhang S., Retroformer: pushing the limits of end-to-end retrosynthesis transformer. *Proceedings. of. the. 39th. International. Conference. on. Machine. Learning,. PMLR*, **22475** (2022).

[122] Fang L., Li J., Zhao M., Tan L. and Lou J.-G., Single-step retrosynthesis prediction by leveraging commonly preserved substructures. *Nat. Commun.*, **14** (2023), 2446.

[123] Schwaller P., Hoover B., Reymond J.-L., Strobelt H. and Laino T., Extraction of organic chemistry grammar from unsupervised learning of chemical reactions. *Sci. Adv.*, **7** (2021), eabe4166.

[124] Ucak U.V., Kang T., Ko J. and Lee J., Substructure-based neural machine translation for retrosynthetic prediction. *J. Cheminf.*, **13** (2021), 4.

[125] Zhang B., Lin J., Du L. and Zhang L., Harnessing data augmentation and normalization preprocessing to improve the performance of chemical reaction predictions of data-driven model. *Polymers*, **15** (2023), 2224.

[126] Zhong Z., Song J., Feng Z., Liu T., Jia L. and Yao S., et al., Root-aligned smiles: a tight representation for chemical reaction prediction. *Chem. Sci.*, **13** (2022), 9023–9034.

[127] Chen B., Shen T., Jaakkola T.S. and Barzilay R., Learning to make generalizable and diverse predictions for retrosynthesis 2019. .

[128] Toniato A., Vaucher A.C., Schwaller P. and Laino T., Enhancing diversity in language based models for single-step retrosynthesis. *Digital. Discovery.*, **2** (2023), 489–501.

[129] Kim E., Lee D., Kwon Y., Park M.S. and Choi Y.-S., Valid, plausible, and diverse retrosynthesis using tied two-way transformers with latent variables. *J. Chem. Inf. Model.*, **61** (2021), 123–133.

[130] Irwin R., Dimitriadis S., He J. and Bjerrum E.J., Chemformer: a pre-trained transformer for computational chemistry. *Mach. Learn:. Sci. Technol.*, **3** (2022), 015022.

[131] Bai R., Zhang C., Wang L., Yao C., Ge J. and Duan H., Transfer learning: making retrosynthetic predictions based on a small chemical reaction dataset scale to a new level. *Molecules.*, **25** (2020), 2357.

[132] Andronov M., Voinarovska V., Andronova N., Wand M., Clevert D.-A. and Schmidhuber J., Reagent prediction with a molecular transformer improves reaction data quality. *Chem. Sci.*, **14** (2023), 3235–3246.

[133] Schreck J.S., Coley C.W. and Bishop K.J.M., Learning retrosynthetic planning through simulated experience. *ACS Cent. Sci.*, **5** (2019), 970–981.

[134] Guo J., Yu C., Li K., Zhang Y., Wang G. and Li S., et al., Retrosynthesis zero: self-improving global synthesis planning using reinforcement learning. *J. Chem. Theory. Comput.*, **20** (2024), 4921–4938.

[135] Mao K., Xiao X., Xu T., Rong Y., Huang J. and Zhao P., Molecular graph enhanced transformer for retrosynthesis prediction. *Neurocomputing.*, **457** (2021), 193–202.

[136] Sun R., Dai H., Li L., Kearnes S. and Dai B., Towards understanding retrosynthesis by energy-based models. *Advances. in. Neural. Information. Processing. Systems,. vol. 34,. Curran. Associates,. Inc.*, **10186** (2021).

[137] Tu Z. and Coley C.W., Permutation invariant graph-to-sequence model for template-free retrosynthesis and reaction prediction. *J. Chem. Inf. Model.*, **62** (2022), 3503–3513.

[138] Liu C.-H., Korablyov M., Jastrzębski S., Włodarczyk-Pruszyński P., Bengio Y. and Segler M., Retrognn: fast estimation of synthesizability for virtual screening and de novo design by learning from slow retrosynthesis software. *J. Chem. Inf. Model.*, **62** (2022), 2293–2300.

[139] Sacha M., Błaż M., Byrski P., Dąbrowski-Tumański P., Chromiński M. and Loska R., et al., Molecule edit graph attention network: modeling chemical reactions as sequences of graph edits. *J. Chem. Inf. Model.*, **61** (2021), 3273–3284.

[140] Thakkar A., Vaucher A.C., Byekwaso A., Schwaller P., Toniato A. and Laino T., Unbiasing retrosynthesis language models with disconnection prompts. *ACS Cent. Sci.*, **9** (2023), 1488–1498.

[141] Wang Y., Pang C., Wang Y., Jin J., Zhang J. and Zeng X., et al., Retrosynthesis prediction with an interpretable deep-learning framework based on molecular assembly tasks. *Nat. Commun.*, **14** (2023), 6155.

[142] Han P., Zhao P., Lu C., Huang J., Wu J. and Shang S., et al., Gnn-retro: retrosynthetic planning with graph neural networks. *Proc. AAAI Conf. Artif. Intell.*, **36** (2022), 4014–4021.

[143] Jiang Y., Wei Y., Wu F., Huang Z., Kuang K. and Wang Z., Learning chemical rules of retrosynthesis with pre-training. *Proc. AAAI Conf. Artif. Intell.*, **37** (2023), 5113–5121.

[144] Liu S., Tu Z., Xu M., Zhang Z., Lin L. and Ying R., et al., Fusionretro: molecule representation fusion via in-context learning for retrosynthetic planning. *Proceedings. of. the. 40th. International. Conference. on. Machine. Learning,. PMLR*, **22028** (2023).

[145] Segler M.H.S., Preuss M. and Waller M.P., Planning chemical syntheses with deep neural networks and symbolic AI. *Nature*, **555** (2018), 604–610.

[146] Lin K., Xu Y., Pei J. and Lai L., Automatic retrosynthetic route planning using template-free models. *Chem. Sci.*, **11** (2020), 3355–3364.

[147] Hong S., Zhuo H.H., Jin K., Shao G. and Zhou Z., Retrosynthetic planning with experience-guided monte carlo tree search. *Commun. Chem.*, **6** (2023), 120.

[148] Latendresse M., Malerich J.P., Herson J., Krummenacker M., Szeto J. and Vu V.-A., et al., Synroute: A retrosynthetic planning software. *J. Chem. Inf. Model.*, **63** (2023), 5484–5495.

[149] Chen B., Li C., Dai H. and Song L., Retro*: learning retrosynthetic planning with neural guided a* search. *Proceedings. of. the. 37th. International. Conference. on. Machine. Learning,. PMLR*, **1608** (2020).

[150] Mikulak-Klucznik B., Gołębiowska P., Bayly A.A., Popik O., Klucznik T. and Szymkuć S., et al., Computational planning of the synthesis of complex natural products. *Nature*, **588** (2020), 83–88.

[151] Klucznik T., Mikulak-Klucznik B., McCormack M.P., Lima H., Szymkuć S. and Bhowmick M., et al., Efficient syntheses of diverse, medicinally relevant targets planned by computer and executed in the laboratory. *Chem.*, **4** (2018), 522–532.

[152] Lin M.H., Tu Z. and Coley C.W., Improving the performance of models for one-step retrosynthesis through re-ranking. *J. Cheminf.*, **14** (2022), 15.

[153] Li J., Fang L. and Lou J.-G., Retroranker: leveraging reaction changes to improve retrosynthesis prediction through re-ranking. *J. Cheminf.*, **15** (2023), 58.

[154] Jeong J., Lee N., Shin Y. and Shin D., Intelligent generation of optimal synthetic pathways based on knowledge graph inference and retrosynthetic predictions using reaction big data. *J. Taiwan. Inst. Chem. Eng.*, **130** (2022), 103982.

[155] Nicolaou C.A., Watson I.A., LeMasters M., Masquelin T. and Wang J., Context aware data-driven retrosynthetic analysis. *J. Chem. Inf. Model.*, **60** (2020), 2728–2738.

[156] Yan C., Ding Q., Zhao P., Zheng S., Yang J. and Yu Y., et al., Retroxpert: decompose retrosynthesis prediction like A chemist. *Chemistry.*, (2020).

[157] Wang X., Li Y., Qiu J., Chen G., Liu H. and Liao B., et al., Retroprime: A diverse, plausible and transformer-based method for single-step retrosynthesis predictions. *Chemical. Engineering. Journal.*, **420** (2021), 129845.

[158] Somnath V.R., Bunne C., Coley C., Krause A. and Barzilay R., Learning graph models for retrosynthesis prediction. *Advances. in. Neural. Information. Processing. Systems,. vol. 34,. Curran. Associates,. Inc.*, **9405** (2021).

[159] Ishida S., Terayama K., Kojima R., Takasu K. and Okuno Y., Ai-driven synthetic route design incorporated with retrosynthesis knowledge. *J. Chem. Inf. Model.*, **62** (2022), 1357–1367.

[160] Zhang B., Zhang X., Du W., Song Z., Zhang G. and Zhang G., et al., Chemistry-informed molecular graph as reaction descriptor for machine-learned retrosynthesis planning. *Proc. Natl. Acad. Sci.*, **119** (2022), e2212711119.

[161] Lin Z., Yin S., Shi L., Zhou W. and Zhang Y.J., G2gt: retrosynthesis prediction with graph-to-graph attention neural network and self-training. *J. Chem. Inf. Model.*, **63** (2023), 1894–1905.

[162] Zhong W., Yang Z. and Chen C.Y.-C., Retrosynthesis prediction using an end-to-end graph generative architecture for molecular graph editing. *Nat. Commun.*, **14** (2023), 3009.

[163] Štrumbelj E. and Kononenko I., Explaining prediction models and individual predictions with feature contributions. *Knowl. Inf. Syst.*, **41** (2014), 647–665.

[164] Ribeiro M.T., Singh S. and Guestrin C., “Why should I trust you?”: explaining the predictions of any classifier. *Proceedings. of. the. 22nd. ACM SIGKDD International. Conference. on. Knowledge. Discovery. and. Data. Mining,. New. York,. NY,. USA:. Association. for. Computing. Machinery.*, **1135** (2016).

[165] Slack D., Hilgard S., Jia E., Singh S. and Lakkaraju H., Fooling LIME and shap: adversarial attacks on post hoc explanation methods. *Proceedings. of. the. AAAI/ACM. Conference. on. AI,. Ethics,. and. Society,. New. York,. NY,. USA:. Association. for. Computing. Machinery.*, **180** (2020).

[166] Ribeiro M.T., Singh S. and Guestrin C., Anchors: high-precision model-agnostic explanations. *Proc. AAAI Conf. Artif. Intell.*, **32** (2018).

[167] Luong M.-T., Pham H. and Manning C.D., Effective approaches to attention-based neural machine translation. *arXiv*, (2015).

[168] Vaswani A., Shazeer N., Parmar N., Uszkoreit J., Jones L. and Gomez A.N., et al., Attention is all you need. *Advances. in. Neural. Information. Processing. Systems,. vol. 30,. Curran. Associates,. Inc.*, (2017).

[169] Houben C. and Lapkin A.A., Automatic discovery and optimization of chemical processes. *Current. Opinion. in. Chemical. Engineering.*, **9** (2015), 1–7.

[170] Gromski P.S., Granda J.M. and Cronin L., Universal chemical synthesis and discovery with ‘the chemputer. ’. *Trends. Chem.*, **2** (2020), 4–12.

[171] Peplow M., Organic synthesis: the robo-chemist. *Nature*, **512** (2014), 20–22.