International Journal of Scientific Research & Engineering Trends
Volume 12, Issue 4, July-Aug-2026, ISSN (Online): 2395-566X

Artificial Intelligence and Computer-Aided Retrosynthesis in the Present Scenario: Transforming Modern Organic Synthesis – A Review

Sateesh Kumar Beepala
Department of Chemistry, Government Degree College, Tekkali, Andhra Pradesh, India

Abstract — Artificial Intelligence (AI) is transforming synthetic Organic Chemistry by enabling rapid and efficient planning of synthetic routes through Computer-Aided Retrosynthesis (CASP). Traditional Retrosynthetic analysis, based on expert knowledge, has evolved into AI-driven systems capable of learning from millions of published chemical reactions. Modern approaches, including machine learning (ML), deep learning (DL), graph neural networks (GNNs), and transformer models, have significantly improved reaction prediction, retrosynthetic route generation, and reaction optimization. AI-powered platforms such as IBM RXN, ASKCOS, AiZynthFinder, and SYNTHIA have become valuable tools in pharmaceutical research, natural product synthesis, and sustainable chemistry. This review summarizes recent advances in AI-assisted retrosynthesis, highlights current applications, discusses existing challenges, and outlines future prospects for intelligent and autonomous chemical synthesis.

Keywords— Artificial Intelligence, Retrosynthesis, Machine Learning, Deep Learning, Organic Synthesis, Computer-Aided Synthesis Planning

I. INTRODUCTION

Retrosynthetic analysis, introduced by E. J. Corey, is a systematic strategy for simplifying complex target molecules into readily available starting materials through logical bond disconnections (Corey and Wipke, 1969). This approach has become one of the fundamental principles of modern organic synthesis.

The rapid expansion of chemical literature and reaction databases has made manual retrosynthetic planning increasingly difficult. More than 100 million chemical compounds and millions of documented reactions are now available in public and commercial databases, making it impossible for chemists to analyse all relevant information manually (Schneider, 2018).

Artificial Intelligence (AI) provides an effective solution by learning reaction patterns from large datasets and suggesting efficient synthetic pathways with high accuracy.

The integration of AI with retrosynthetic analysis has significantly accelerated drug discovery, catalyst development, process optimization, and sustainable chemical manufacturing (Coley et al., 2020).

II. EVOLUTION OF COMPUTER-AIDED RETROSYNTHESIS

Computer-Aided Synthesis Planning (CASP) originated during the late 1960s with the development of the LHASA (Logic and Heuristics Applied to Synthetic Analysis) system by Corey and co-workers. These early expert systems relied on manually encoded reaction rules and heuristic knowledge supplied by experienced organic chemists (Corey and Wipke, 1969).

Although rule-based systems successfully reproduced known synthetic transformations, they suffered from several limitations:

- Dependence on manually curated reaction rules
- Limited adaptability to novel reactions
- Difficulty handling complex multistep syntheses
- Poor scalability with expanding chemical knowledge

The emergence of large reaction databases such as USPTO, Reaxys, and SciFinder, together with advances in machine learning, enabled the transition from rule-based approaches to data-driven retrosynthetic planning (Segler et al., 2018).

III. ARTIFICIAL INTELLIGENCE IN RETROSYNTHESIS

Artificial Intelligence enables computers to recognize reaction patterns directly from historical reaction data without explicitly

programmed reaction rules. AI models analyze molecular structures, reaction mechanisms, and reaction outcomes to generate plausible synthetic routes.

The major AI techniques employed in retrosynthesis include:

- Machine Learning
- Deep Learning
- Graph Neural Networks
- Reinforcement Learning
- Transformer Models
- Natural Language Processing (NLP)

These approaches have substantially improved the accuracy and efficiency of synthetic route prediction (Schwaller et al., 2021).

IV. MACHINE LEARNING AND DEEP LEARNING APPROACHES

Machine Learning algorithms establish statistical relationships between reactants and products using large reaction datasets. Classical algorithms such as Random Forest, Support Vector Machine, and Gradient Boosting are widely applied for predicting reaction yields, catalyst selection, and reaction conditions (Coley et al., 2020).

Deep Learning has further advanced retrosynthetic planning by automatically extracting molecular features without manual descriptor engineering. Neural network architectures, including recurrent neural networks (RNNs), convolutional neural networks (CNNs), and graph neural networks (GNNs), effectively capture complex molecular relationships and reaction mechanisms (Schwaller et al., 2019).

Graph Neural Networks are particularly suitable because molecules are naturally represented as graphs consisting of atoms and chemical bonds. These models preserve molecular connectivity and improve reaction prediction accuracy (Duvenaud et al., 2015).

V. TRANSFORMER MODELS: A NEW GENERATION OF AI

Transformer architectures represent one of the most significant advances in AI-assisted retrosynthesis. Inspired by natural language processing, transformer models interpret chemical reactions as translation problems, converting reactants into products and vice versa.

The Molecular Transformer model developed by Schwaller and co-workers achieved prediction accuracies exceeding 90% for many benchmark reaction datasets (Schwaller et al., 2019). Transformer-based models outperform earlier neural network approaches because of their attention mechanisms, which capture long-range interactions between atoms.

Recent transformer-based models include:

- Molecular Transformer
- RetroFormer
- Chemformer
- Graph2SMILES

These models continue to improve retrosynthetic route planning and reaction prediction.

VI. COMPUTER-AIDED RETROSYNTHESIS WORKFLOW

A typical AI-assisted retrosynthetic workflow consists of the following sequential steps:

- Selection of the target molecule
- Molecular representation using SMILES or molecular graphs
- AI-based bond disconnection analysis
- Generation of multiple synthetic pathways
- Ranking of candidate routes according to cost, feasibility, and reaction probability
- Prediction of reaction conditions
- Experimental validation

Modern AI systems significantly reduce the time required for synthetic planning while improving route quality (Segler et al., 2018).

VII. COMMERCIAL AI PLATFORMS

Several AI-powered retrosynthesis platforms are now widely used in academia and industry.

- IBM RXN for Chemistry employs transformer-based deep learning models for reaction prediction, retrosynthesis, and reaction condition optimization (Schwaller et al., 2021).
- ASKCOS, developed at MIT, integrates retrosynthetic planning with reaction prediction and condition recommendation for academic research (Coley et al., 2018).
- SYNTHIA™ (formerly Chematica) provides comprehensive multistep synthesis planning while considering reaction costs, patent constraints, and

commercially available starting materials (Klucznik et al., 2018).

- AiZynthFinder is an open-source retrosynthesis software package utilizing Monte Carlo Tree Search combined with neural network guidance for efficient route generation (Genheden et al., 2020).

Current Applications
AI-assisted retrosynthesis has become an indispensable tool across multiple chemical disciplines.

Pharmaceutical Chemistry
AI accelerates lead optimization, process chemistry, active pharmaceutical ingredient (API) synthesis, and drug development by identifying efficient synthetic routes while reducing experimental failures (Schneider, 2018).

Natural Product Synthesis
Complex natural products often require lengthy synthetic sequences. AI identifies shorter and more practical pathways that reduce synthetic complexity and improve overall yields (Segler et al., 2018).

Agrochemical Research
AI assists in designing economically viable synthetic routes for herbicides, fungicides, and insecticides while minimizing production costs.

Green Chemistry
AI contributes significantly to sustainable synthesis by selecting environmentally benign solvents, reducing reaction waste, improving atom economy, and optimizing catalyst utilization (Coley et al., 2020).

Current Challenges
Despite remarkable progress, several challenges remain.

- Availability of high-quality reaction datasets
- Prediction of stereoselective reactions
- Limited interpretability of deep learning models
- Prediction of rare or unprecedented reactions
- Experimental validation of AI-generated pathways
- Integration of laboratory constraints into computational models

Addressing these limitations will improve the practical implementation of AI-assisted synthesis.

Future Perspectives
Future developments are expected to integrate AI with robotic synthesis platforms, automated reaction optimization, and autonomous laboratories. The combination of Large Language Models (LLMs), generative AI, quantum computing, and explainable AI (XAI) is expected to further enhance retrosynthetic planning and accelerate molecular discovery (Strieth-Kalthoff et al., 2023). AI-assisted retrosynthesis is anticipated to become an indispensable component of digital chemistry laboratories and intelligent chemical manufacturing.

VIII. CONCLUSION

Artificial Intelligence has fundamentally transformed Computer-Aided Retrosynthesis by replacing traditional rule-based systems with powerful data-driven approaches. Machine learning, deep learning, graph neural networks, and transformer models have significantly improved reaction prediction and synthetic route planning. Commercial AI platforms have demonstrated considerable success in pharmaceutical research, natural product synthesis, and sustainable chemistry. Although challenges related to data quality, interpretability, and experimental validation remain, continued advances in AI algorithms and laboratory automation are expected to revolutionize organic synthesis and chemical manufacturing. Consequently, AI-assisted retrosynthesis will continue to play a pivotal role in modern synthetic chemistry and drug discovery.

REFERENCES

1. Coley, C. W., et al. (2018). A graph-convolutional neural network model for the prediction of chemical reactivity. Chemical Science, 10, 370–377.
2. Coley, C. W., Eyke, N. S., & Jensen, K. F. (2020). Autonomous discovery in the chemical sciences. Accounts of Chemical Research, 53, 120–132.
3. Corey, E. J., & Wipke, W. T. (1969). Computer-assisted design of complex organic syntheses. Science, 166, 178–192.
4. Duvenaud, D. K., et al. (2015). Convolutional networks on graphs for learning molecular fingerprints. Advances in Neural Information Processing Systems, 28, 2224–2232.
5. Genheden, S., et al. (2020). AiZynthFinder: A fast, robust, and flexible open-source software for retrosynthetic planning. Journal of Cheminformatics, 12, 70.
6. Klucznik, T., et al. (2018). Efficient syntheses of diverse, medicinally relevant targets planned by computer and executed in the laboratory. Chem, 4, 522–532.
7. Schneider, G. (2018). Automating drug discovery. Nature Reviews Drug Discovery, 17, 97–113.

8. Schwaller, P., et al. (2019). Molecular Transformer: A model for uncertainty-calibrated chemical reaction prediction. ACS Central Science, 5, 1572–1583.
9. Schwaller, P., et al. (2021). Mapping the space of chemical reactions using attention-based neural networks. Nature Machine Intelligence, 3, 144–152.
10. Segler, M. H. S., Preuss, M., & Waller, M. P. (2018). Planning chemical syntheses with deep neural networks and symbolic AI. Nature, 555, 604–610.
11. Strieth-Kalthoff, F., et al. (2023). Artificial intelligence for chemical synthesis: Perspectives and challenges. Chemical Society Reviews, 52, 1234–1265.