# Supporting Information

**Transforming Drug Discovery: The Role of AI in Chemical Synthesis Planning**

Madhavi Kumaria, Pradnya Kamblea, M. N. V. Ravi Kumarb, Prabha Garga\*

aDepartment of Pharmacoinformatics, National Institute of Pharmaceutical Education and Research (NIPER) S.A.S. Nagar, Sector 67, S.A.S. Nagar (Mohali)-160062, Punjab (India)

bDivision of Translational Science and Medicine, College of Community Health Sciences, and Convergent Bioscience and Medicine, The University of Alabama, Tuscaloosa 35401 (USA)

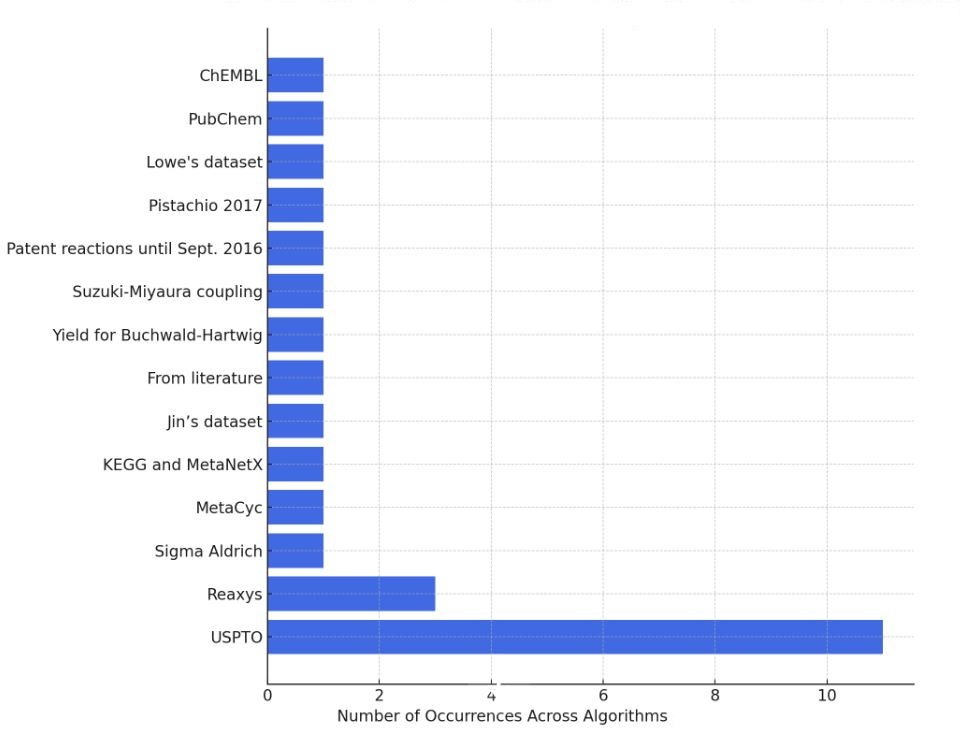
# **Researches and software developed for synthesis planning by ML**

Recent advancements in ML have significantly impacted the field of synthesis planning in chemistry, which have led to the development of sophisticated software tools. These tools are designed to assist chemists in mapping out the most efficient and feasible pathways for synthesising complex molecules. By utilising algorithms that can process vast datasets and identify patterns and relationships within chemical structures and reactions, these software programs significantly enhance the speed and accuracy of synthesis planning.

Moreover, the development of synthesis-planning programs increasingly relies on a combination of human insight and machine learning. This hybrid approach leverages the strengths of both human expertise in chemistry and the computational power of ML algorithms. Such programs aim to replicate the decision-making process of synthetic chemists, thereby facilitating more efficient and innovative approaches to chemical synthesis. These advancements reflect a broader trend in the pharmaceutical and chemical industries, where high-throughput experimentation and data-driven methods are revolutionising traditional practices. There are many software and tools developed for synthesis planning (Table 1)

**Table 1.** Various software and tools developed for synthesis planning by ML

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| **Software/Tool** | **Algorithm** | **Datasets Used** | **Approach** | **Ref** | **GitHub Code** | **Accuracy** | **Stereochemistry considered** |
| AiZynthFinder | Monte Carlo tree search | USPTO  dataset | Retrosynthesis Prediction | [1] | https://github.com/MolecularAI/aizynthfinder |  |  |
| ASKCOS | Monte Carlo tree search | Reaxys, Sigma Aldrich | Reaction Prediction, Retrosynthesis Prediction | [2] |  |  | Yes |
| AutoSynRoute | MCTS | USPTO\_50K, USPTO\_MIT | retrosynthetic prediction | [3] | http://https//github.com/PKUMDL-AI/AutoSynRoute | 63.0% | Yes |
| 3N-MCTS | MCTS | Reaxys | Retrosynthesis prediction | [4] |  |  | Yes |
| BioNavi-NP | AND-OR tree-based | MetaCyc, KEGG and MetaNetX | Bio-Retrosynthesis Prediction | [5] | https://github.com/prokia/BioNavi-NP | 60.6% | Yes |
| NeuralSym | Expert system,  Logistic Regression,  Highway Network,  FC512 ELU | Reaxys | Reaction Prediction, Retrosynthesis | [6] | https://github.com/linminhtoo/neuralsym | 97% (Reaction prediction),  95% (Retrosynthesis) |  |
| ChemAI | kNN, SVM, Gradient boosting, Deep Neural Networks (DNN), graph neural networks |  | Reaction Prediction, Retrosynthesis | https://ai.chemdx.org/ | Commercial |  |  |
| SCROP | transformer neural networks | USPTO,  Jin’s data set | Retrosynthesis Prediction | [7] | https://github.com/Jh-SYSU/SCROP | 59.0% (USPTO),  41.5% (Jin’s data set) |  |
| RetroXpert | Reactant Generation Network | USPTO | Retrosynthesis Prediction | [8] | https://github.com/uta-smile/RetroXpert | 70.4% (with reaction type),  65.5% (without reaction type) |  |
| DeepReac+ | GNN-based deep learning | From literature  Dataset A[9]  Dataset B [10]  Dataset C[11] | Chemical reaction outcome prediction,  Identification of optimal reaction conditions | [12] | https://github.com/bm2-lab/DeepReac |  |  |
| GraphRXN | Graph-sum, Graph-concat, Yield-BERT, | Yield for Buchwald-Hartwig, Suzuki-Miyaura coupling, reaction, Stereo-selectivity for asymmetric N, S-acetal formation reaction | Reaction Prediction | [13, 14] | https://github.com/jidushanbojue/GraphRXN | 0.713 |  |
| IBM RXN for Chemistry | Deep Neural Networks (DNN) |  | Reaction Prediction | https://rxn.res.ibm.com/ | Commercial |  | Yes |
| ICSYNTH |  |  |  | [15] |  |  | Yes |
| MechRetro | Graph Transformer architecture | USPTO-50K | Retrosynthetic Prediction and Pathway Plan | [16] | https://github.com/Wanyinee/MechR |  |  |
| Molecular Transformer | Autoregressive encoder–decoder model | USPTO MIT, USPTO LEF, USPTO STERE, Patent reactions until Sept. 2016, Pistachio 2017 | Reaction Prediction | [17] |  | 90.40% | Yes |
| Neural seq2seq model | LSTM, BLSTM | Jin's USPTO dataset, Lowe's dataset | Reaction Prediction | [18] |  | 80.30%, 0.654 |  |
| ReactionMiner |  | KEGG database | Reaction Prediction, Pathway Prediction | [19] | https://github.com/RamanLab/ReactionMiner |  |  |
| ReactionPredictor | ANNs | EP, new Reaction Explorer, graduate texts, RP | Reaction Prediction | [20] |  | >77% |  |
| RetroExplainer |  | USPTO50K, USPTO-MIT | Retrosynthesis Prediction |  | https://github.com/wangyu-sd/RetroExplainer |  |  |
| RetroPath2.0 |  |  | Reaction Prediction | [21] | https://github.com/brsynth/rp2paths |  |  |
| RetroPrime |  | USPTO-50 K, USPTO-full | Retrosynthesis Prediction | [22] | https://github.com/wangxr0526/RetroPrime | 51 -65%(USPTO-50 K)  44.1%(USPTO-full) |  |
| RetroPathRL | Monte Carlo Tree Search |  | Retrosynthesis Prediction | [23] | https://github.com/brsynth/RetroPathRL |  |  |
| RetroTRAE | Transformer-based AE model | USPTO, PubChem, ChEMBL | Retrosynthesis Prediction | [24] | https://github.com/knu-lcbc/RetroTRAE | 58.30% |  |
| Synthia (Chematica) |  |  | Reaction Prediction, Retrosynthesis | [25] | Commercial |  | Yes |
| Graph2SMILES |  | USPTO,  CAS | Reaction Prediction, Retrosynthesis | [26] | https://github.com/coleygroup/Graph2SMILES |  | Yes |
| SEMG-MIGNN | SEMG, MIGNN | Doyle’s Dataset[9], Denmark’s Dataset[27] | Predicting reaction yields | [28] | https://github.com/Shuwen-Li/SEMG-MIGNN |  |  |
| Chemist-X | RAG, SCL | PubChem, ChemSpi-der | Reaction condition recommendation | [29] |  |  |  |



**Figure S1**: Most used dataset across different algorithms

# References

1. Genheden, S., et al., *AiZynthFinder: a fast, robust and flexible open-source software for retrosynthetic planning.* J Cheminform, 2020. **12**(1): p. 70.

2. Gallou, F., R. Robinson, and M. Parmentier, *Computer Aided Synthesis Planning: ASKCOS.* Journal of medicinal chemistry, 2020.

3. Lin, K., et al., *Automatic retrosynthetic route planning using template-free models.* Chem Sci, 2020. **11**(12): p. 3355-3364.

4. Segler, M.H.S., M. Preuss, and M.P. Waller, *Planning chemical syntheses with deep neural networks and symbolic AI.* Nature, 2018. **555**(7698): p. 604-610.

5. Zheng, S., et al., *Deep learning driven biosynthetic pathways navigation for natural products with BioNavi-NP.* Nature Communications, 2022. **13**(1): p. 3342.

6. Segler, M.H.S. and M.P. Waller, *Neural-Symbolic Machine Learning for Retrosynthesis and Reaction Prediction.* Chemistry, 2017. **23**(25): p. 5966-5971.

7. Zheng, S., et al., *Predicting Retrosynthetic Reactions Using Self-Corrected Transformer Neural Networks.* J Chem Inf Model, 2020. **60**(1): p. 47-55.

8. Yan, C., et al., *Retroxpert: Decompose retrosynthesis prediction like a chemist.* Advances in Neural Information Processing Systems, 2020. **33**: p. 11248-11258.

9. Ahneman, D.T., et al., *Predicting reaction performance in C-N cross-coupling using machine learning.* Science, 2018. **360**(6385): p. 186-190.

10. Perera, D., et al., *A platform for automated nanomole-scale reaction screening and micromole-scale synthesis in flow.* Science, 2018. **359**(6374): p. 429-434.

11. Zahrt, A.F., et al., *Prediction of higher-selectivity catalysts by computer-driven workflow and machine learning.* Science, 2019. **363**(6424).

12. Gong, Y., et al., *DeepReac+: deep active learning for quantitative modeling of organic chemical reactions.* Chem Sci, 2021. **12**(43): p. 14459-14472.

13. Li, B., et al., *GraphRXN: A Novel Representation for Reaction Prediction.* 2022.

14. Li, B., et al., *A deep learning framework for accurate reaction prediction and its application on high-throughput experimentation data.* J Cheminform, 2023. **15**(1): p. 72.

15. Bøgevig, A., et al., *Route design in the 21st century: the IC SYNTH software tool as an idea generator for synthesis prediction.* Organic Process Research & Development, 2015. **19**(2): p. 357-368.

16. Wang, Y., et al., *MechRetro is a chemical-mechanism-driven graph learning framework for interpretable retrosynthesis prediction and pathway planning.* arXiv preprint arXiv:2210.02630, 2022.

17. Schwaller, P., et al., *Molecular transformer: a model for uncertainty-calibrated chemical reaction prediction.* ACS central science, 2019. **5**(9): p. 1572-1583.

18. Schwaller, P., et al., *“Found in Translation”: predicting outcomes of complex organic chemistry reactions using neural sequence-to-sequence models.* Chemical science, 2018. **9**(28): p. 6091-6098.

19. Sankar, A., S. Ranu, and K. Raman, *Predicting novel metabolic pathways through subgraph mining.* Bioinformatics, 2017. **33**(24): p. 3955-3963.

20. Kayala, M.A. and P. Baldi, *ReactionPredictor: prediction of complex chemical reactions at the mechanistic level using machine learning.* Journal of chemical information and modeling, 2012. **52**(10): p. 2526-2540.

21. Delépine, B., et al., *RetroPath2.0: A retrosynthesis workflow for metabolic engineers.* Metab Eng, 2018. **45**: p. 158-170.

22. Wang, X., et al., *RetroPrime: A Diverse, plausible and Transformer-based method for Single-Step retrosynthesis predictions.* Chemical Engineering Journal, 2021. **420**: p. 129845.

23. Jin, W., et al., *Predicting organic reaction outcomes with weisfeiler-lehman network.* Advances in neural information processing systems, 2017. **30**.

24. Ucak, U.V., et al., *Retrosynthetic reaction pathway prediction through neural machine translation of atomic environments.* Nat Commun, 2022. **13**(1): p. 1186.

25. Grzybowski, B.A., et al., *Chematica: A Story of Computer Code That Started to Think like a Chemist.* Chem, 2018. **4**: p. 390-398.

26. Tu, Z. and C.W. Coley, *Permutation Invariant Graph-to-Sequence Model for Template-Free Retrosynthesis and Reaction Prediction.* J Chem Inf Model, 2022. **62**(15): p. 3503-3513.

27. Zahrt, A.F., et al., *Prediction of higher-selectivity catalysts by computer-driven workflow and machine learning.* Science, 2019. **363**(6424): p. eaau5631.

28. Li, S.W., et al., *Reaction performance prediction with an extrapolative and interpretable graph model based on chemical knowledge.* Nat Commun, 2023. **14**(1): p. 3569.

29. Chen, K., *Chemist-X: Large Language Model-empowered Agent for Reaction Condition Recommendation in Chemical Synthesis.* Preprint at, 2024.