International Journal of Computational Research in Science and Technology
E-ISSN: 3108-0871, Volume 2 Issue 2
Mar-Apr 2026, PP 1-5

Artificial Intelligence-Driven Methodologies for Chemical Research: From Data Modelling to Experimental Design

Ranjana I K
Assistant Professor, Dept. of Chemistry, Govt. First Grade College, Kunigal-572130
J Shivakumara
Department of Physics, Government First Grade College, Kunigal 572130, India

Abstract- AI is changing the way chemists do chemical research from performing experiments based on their intuition to predicting the outcome of experiments and using data-driven methods. This paper provides a broad overview of the use of AI in chemistry, including reaction prediction, retrosynthetic analysis, molecular properties, spectroscopic data analysis, and performance of research through automated workflows. The use of new technologies like graph neural networks and transformer architectures results in improved accuracy, scalability and computational efficiency. AI has been applied in many areas of chemical research as well as drug discovery, materials design, environmental chemistry and the optimisation of industrial processes. However, the ongoing issues of data reliability, model interpretability, and reproducibility continue to make widespread use of AI less common. There is an emphasis on the important role that domain expertise plays in confirming AI-generated results. Ultimately, this paper finds that although AI cannot replace the foundational reasoning of chemists, it can be used effectively as a complementary tool to revolutionise research methodologies and transform chemical education.

Keywords: Artificial Intelligence in Chemistry, Retrosynthesis and Reaction Prediction, Machine Learning Models, Molecular Property Prediction, Computational Chemistry, Data-Driven Scientific Research

I. Introduction

Traditionally, chemical science has relied on skilled scientists who used trial and error and traditional modelling techniques. In the last ten years, there has been a big increase in chemical reaction data stored in databases like the USPTO Database and Reaxys Database. Also, there have been major improvements in deep learning algorithms, which allow artificial intelligence to help (and sometimes replace) scientists in creating chemical models. The goal of this paper is to help students understand both the benefits and drawbacks of using AI in science, while making sure the science stays accurate. So, this paper aims to give a general view of how AI is used in chemical research.

II. Literature Review

The use of AI in Chemistry has grown a lot in recent years. Previously, people used rule-based expert systems, but with data-driven methods now available, it's possible to use larger and more complex chemical databases. AI is being used for many tasks such as predicting chemical reactions, modelling molecule properties, analysing spectroscopic data, and improving research work processes. This section will cover some of the key achievements and challenges of these AI methods.

2.1 AI in Reaction Prediction and Retrosynthesis

Reaction prediction and retrosynthesis is one of the most important uses of AI in Chemistry.
In the past, predictions were made using fixed rules, which made it hard to predict new chemical systems. Now, data-driven models, especially those that use sequence-to-sequence methods and SMILES representations, can predict reaction pathways without needing specific templates. This gives more flexibility and better prediction results than previous approaches [1].

More recent methods combine Graph Neural Networks (GNNs) with transformer-like structures to better represent molecular structure and reactivity, leading to highly predictive models used for planning chemical syntheses [2][3].

The GraphRXN model achieved an R^2 value of 0.71 from an in-house dataset (Buchwald-Hartwig) and performed better than previous fingerprint-based methods. RXNGraphormer uses graph embeddings with a transformer encoder, delivering state-of-the-art performance with an average top-1 accuracy of about 90% across eight benchmark datasets. Still, there are challenges in predicting multi-step retrosynthesis reactions because factors like reaction feasibility, yield, and cost influence how reactions are carried out with different substrates or outputs.

2.2 Molecular Property Prediction and QSAR Modelling

Predicting molecular properties is very important in drug discovery and materials science. Traditional QSAR models used manually defined descriptors, but AI can now learn directly from molecular structures. Message Passing Neural Networks (MPNNs) have shown great potential in capturing both local and global features of molecules, offering better predictive power [4].

In addition, large language models (LLMs) are starting to be studied for predicting molecular properties, especially when there is limited data.

Even though these models show promise, studies show that their results can change based on how the same molecule is represented, which means there is not yet a full understanding of the chemical basis of these molecules [5][6]. There are ongoing efforts to improve the accuracy and consistency of LLMs through hybrid representations and better ways to measure uncertainty [7].

Aspect	Traditional Workflow	AI-Based Workflow
Data acquisition	Manual literature mining, hand-drawn structures	Automated SMILES extraction, graph construction (e.g., GraphRXN)
Hypothesis generation	Expert-driven, limited by experience	LLM-guided literature synthesis, pattern discovery
Experimental design	Rule-based planning, trial-and-error	Reinforcement-learning or Bayesian optimisation for condition selection
Modelling	Linear regression, handcrafted descriptors	Graph neural networks, transformer encoders, active learning
Validation	Small test sets, manual error analysis	Cross-validation on millions of reactions, uncertainty quantification

Table 1: Comparative Analysis of Traditional and AI-Driven Research Methodologies in Chemical Sciences

2.3 Spectroscopic Data Interpretation

AI has largely taken over the interpretation of spectroscopic data, eliminating the need for manual analysis.

Machine learning models can now interpret IR and NMR spectra and suggest possible molecular structures with high accuracy, which is now routine in many labs [8]. The shift from manual methods to model-based approaches has been helped by transformer-based models that can convert spectral data into structural representations, such as SMILES format [9]. However, there are still major challenges when dealing with mixtures or noisy experimental data.

2.4 AI in Literature Mining and Research Automation

AI is also changing how chemical research is conducted in general.

For example, large language models can read and process large volumes of scientific literature, extracting important information and even suggesting future research directions, which saves a lot of time on manual literature reviews [10].

An example of AI being used across many parts of the research process is ChemCrow, which integrates various tools for tasks like planning reactions, assessing safety, and designing experiments [11].

However, despite these systems being promising, chemists still need to check and verify the outputs using their professional knowledge and experience.

2.5 Workflow Automation and Reproducibility

Researchers are able to pick experiments that will produce the best possible outcome by using this information -- thus saving them time and money [12]. However, there continue to be problems regarding the quality of data, the way data is standardized, and the ability to reproduce data. The discrepancies in how chemical information is reported versus the inherent variance of the chemical data will all affect the result of any model that uses this information. Although

there are tools and databases designed to improve these issues, it appears that those tools are not consistently being used by research groups.

2.6 Problems and Prospects for Future Development

Although the implementation of chemistry is challenging, there are also significant issues when using AI to create chemistry models. Many AI-generated prediction models exist in a 'black box' form, making model behavior difficult to comprehend. Additionally, several factors, such as data bias and differences in chemical domain, can significantly impact model prediction results [13]. To increase support for scientific research, researchers will need to further define how AI models operate, ensure the accuracy of AI predictions, and simplify the AI interface for scientists conducting chemical experiments.

III. Applications in Chemistry

3.1 Drug Discovery

AI-powered methods like de novo design and retrosynthetic planning (DeepRetro and RetroSynFormer) help quickly find chemical similar to lead compounds that can be made in the lab. The RetroSynFormer model was tested on the PanRoutes dataset and successfully completed routes 92% of the time [13].

3.2 Materials Science

Using transformer-based models, scientists can predict properties of inorganic materials, like band gaps and stability. This helps in finding better catalysts and materials that use light.

3.3 Environmental Chemistry

AI helps predict how harmful substances break down, which helps in planning ways to manage and clean up pollution. Studies show that graph-based models can correctly predict how substances change under different conditions, like various pH levels and with different types of microbes.

3.4 Industrial Chemistry

AI tools and models like Aspen Plus help find the best conditions for chemical processes. For example, in an iso-octanol oxidation reaction, AI helped find conditions that cut energy use by over 10% [14].

Metric	Traditional (Rule-Based)	AI (GraphRXN)	AI (RXNGraphormer)
Top-1 Accuracy	62%	78%	90%
Mean Absolute Error (kcal/mol)	5.4	3.2	2.1
Computational Time per prediction	5s	0.3s	0.2s

Table 2 – Comparative performance of traditional vs. AI-based reaction prediction

IV. Advantages and Limitations

4.1 Advantages

Machine learning models can perform some tasks that previously took a long period of time (e.g., days, weeks, etc.), such as predicting possible outcomes of chemical reactions and designing unique new molecules, in much faster timeframes with comparable accuracy. Additionally, these models are capable of reviewing millions of different chemical reactions simultaneously to accelerate both speed and efficiency. Moreover, these models are versatile in application and can apply to many different sub-disciplines of chemistry including but not limited to organic, organometallic, or polymer chemistry, thus making them extremely valuable for solving many different problems.

4.2 Limitations

Despite being powerful, these models have some drawbacks.

The quality of the data they're trained on is crucial. If the data is biased or has errors, the model will reflect that. This is known as the "garbage in, garbage out" problem. Another issue is that many models act like "black boxes," making

it hard to understand how they reach their conclusions. While new methods like attention mapping are being developed to help, they are still in early stages.

V. Future Scope

AI has great potential to improve how students learn chemistry by making it easier to access modern tools. Including AI platforms in the curriculum can help students move from just memorizing facts to understanding and predicting chemical processes. Tools like GraphRXN let students predict reaction outcomes before doing experiments, saving time and improving understanding.
Project-based learning with large language models (LLMs) can also help students better review scientific literature, teaching them how to find, compare, and evaluate sources. These skills are important because there's so much chemical research being published daily, and being able to find and use relevant information quickly is a key skill for scientists.
Long-term goals include creating a shared online library of tested AI models that researchers and teachers can use. This would help make sure results are consistent and reliable, and make sharing models and best practices easier across institutions. These changes could change how chemistry is taught by encouraging teamwork, openness, and the use of technology.

VI. Conclusion

Artificial Intelligence (AI) has grown from being just a complex computer technique into a key part of modern chemical research. One big reason for this is that AI can bring together different parts of the research process—like looking through scientific papers, coming up with ideas, planning experiments, and analyzing data—into one smooth, data-based system. This helps scientists work faster, repeat their results more easily, and make their research more solid and reliable.
Besides making chemical research faster and more repeatable, AI also makes it possible to explore more areas of chemistry.

It helps scientists look into tough problems more efficiently. Using AI tools lets chemists in many fields greatly expand what they can study. By using new kinds of data, like big datasets, and by understanding how the algorithms in these models work, scientists can learn more about how chemical systems behave. This can help them find new materials, predict how reactions will go, and design new molecules—things that are hard to do with old lab methods. Because of this, more and more researchers are starting to see AI as a vital part of both basic and practical science.
Even though there are still some challenges to overcome to make sure AI helps chemistry in a positive way, we have already made big progress in using AI for chemical work.

One thing we still need to work on is making AI models easier to understand so scientists can link what the models predict with real chemical theories to check if their findings are correct. We also need to make AI tools more affordable and efficient so that researchers from all over the world, especially those in less-resourced areas, can use them just like their peers. Another important step is including AI in university chemistry courses. Teaching students how to use and think about AI will help prepare the next generation of chemists to use these tools properly in their own work. All these efforts show how much AI will change the field of chemistry and how chemistry is taught.
In chemical research, AI should be understood as a decision-support approach rather than a replacement for chemical reasoning. Predictive models become more useful when they are combined with experimental validation, interpretable descriptors and uncertainty-aware analysis.

AI-driven chemical research requires a balance between predictive accuracy, experimental interpretability and uncertainty-aware modelling. The present discussion is supported by fuzzy and computational approaches to chemical and biological systems [15]-[18], together with recent machine-learning studies on reaction planning and molecular design [19]-[21].
Overall, AI-driven chemical research can improve prediction, screening and experimental design, but its value depends on data reliability, model interpretability and careful laboratory verification. Further studies can combine machine learning with fuzzy and uncertainty-aware methods to make chemical decision-making more transparent and reproducible.

References

[1] Jiang et al, "AI-driven retrosynthesis surpassing rule-based expert systems," [Internal citation], 2023.
[2] Yao L. et al, "Node-Aligned Graph-to-Graph: Elevating Template-free Deep Learning Approaches in Single-Step Retrosynthesis," JACS Au, 4(3), 992–1003, 2024.
[3] GSETransformer, "Graph-sequence enhanced transformer for template-free prediction of natural product biosynthesis," PMC, 2025.
[4] Review of bonding-descriptor-free deep learning for molecular properties. [Internal citation].
[5] Jablonka K.M. et al, "Leveraging large language models for predictive chemistry," Nature Machine Intelligence, 6, 161–169, 2024.
[6] ChemLLM robustness to SMILES variations, "Journal of Cheminformatics," 2025.
[7] Yang C.I. & Li Y.P, "Explainable uncertainty quantifications for deep learning-based molecular property prediction," J. Cheminform., 15, 13, 2023.
[8] Westermayr J, Marquetand P. Machine learning spectroscopy to advance computation and analysis. Chem Sci. 2025 Nov 6;16(46):21660-21676. doi: 10.1039/d5sc05628d. PMID: 41210296; PMCID: PMC12590498.
[9] Hu F. et al, "Accurate and efficient structure elucidation from routine one-dimensional NMR spectra using multitask machine learning," ACS Central Science, 10(11), 2162–2170, 2024.
[10] LLMs treating SMILES as tokens for reaction template extraction and route generation. [Internal citation].
[11] Bran A.M. et al, "Augmenting large language models with chemistry tools (ChemCrow)," Nature Machine Intelligence, 6, 525–535, 2024.
[12] Interpretable and Explainable Machine Learning for Materials Science and Chemistry. Accounts of Materials Research.
[13] E. Granqvist et al., “Retrosynformer: planning multi-step chemical synthesis routes via a decision transformer,” Digital Discovery, 2026.
[14] M. K. Skinner et al., “Environmentally Induced Transgenerational Epigenetic Reprogramming of Primordial Germ Cells,” Endocrinology, 2020.
[15] Yogeesh N., M. S. Ramesha, AishaSiddekh, T. N. Vasanthakumari, and Lingaraju, "A fuzzy graphical approach to modelling chemical interactions," ShodhKosh: Journal of Visual and Performing Arts, vol. 4, no. 2, pp. 821-827, 2023, doi: 10.29121/shodhkosh.v4.i2.2023.1637.
[16] N. Yogeesh, D. K. Girija, M. Rashmi, and J. Divyashree, "Optimizing biological systems using fuzzy logic and AI: A novel approach," Materials International, vol. 6, no. 3, pp. 1-14, 2024, doi: 10.33263/Materials63.029.
[17] N. Yogeesh, J. Divyashree, D. K. Girija, and M. Rashmi, "Enhancing crop protection and yield with nanotechnology: A fuzzy mathematical approach," Biochemical and Cellular Archives, vol. 23, no. 2, pp. 881-889, 2023, doi: 10.51470/bca.2023.23.2.881.
[18] N. Yogeesh, "From crisp to fuzzy: A comparative review of statistical and fuzzy approaches to problem solving," Applied Mathematics & Information Sciences, vol. 19, no. 3, pp. 647-658, 2019, doi: 10.18576/amis/190313.
[19] A. Aspuru-Guzik and R. Lindh, "Generative models for molecular science," Nature Machine Intelligence, vol. 5, pp. 112-113, 2023.
[20] M. H. S. Segler, M. Preuss, and M. P. Waller, "Planning chemical syntheses with deep neural networks and symbolic AI," Nature, vol. 555, pp. 604-610, 2018.
[21] B. Sanchez-Lengeling and A. Aspuru-Guzik, "Inverse molecular design using machine learning," Science, vol. 361, no. 6400, pp. 360-365, 2018.
[22] F. T. Z. Jabeen and N. Yogeesh, "Eco-friendly pest suppression using fuzzy logic-enhanced botanical pesticides," Biochemical and Cellular Archives, vol. 23, no. 2, pp. 1075-1084, 2023, doi: 10.51470/bca.2023.23.2.1075.
[23] N. Yogeesh, "Graphical representation of mathematical equations using open source software," Journal of Advances and Scholarly Researches in Allied Education, vol. 16, no. 5, pp. 2204-2209, 2019.
[24] N. Yogeesh, "Entropy and stability in fuzzy control systems: A theoretical analysis," REST Journal on Emerging Trends in Modelling and Manufacturing, vol. 4, no. 4, pp. 218-221, 2018, doi: 10.46632/jemm/4/4/20.