Review Paper:

Exploring Artificial Intelligence in Organic Synthesis

Srivastava A.[1]*, Singh R.[2], Sahai A.[3] and Srivastava N.[4]*
1. Department of Chemistry, GLA University, Mathura, 281406, U.P., INDIA
2. Department of Chemistry, B. N College of Engineering and Technology, Lucknow, U.P., INDIA
3. Department of Computer Science, Galgotias University, Greater Noida, 203201, U.P., INDIA
4. Department of Chemistry, Amity University Uttar Pradesh Lucknow Campus, 226028, U.P., INDIA
*aabhichem@gla.ac.in; nitinsriv5@gmail.com

Abstract

The field of chemistry has seen significant growth with the integration of Artificial Intelligence (AI). This has provided chemists with advanced techniques and tools that make the interpretation of chemical problems much easier and faster. In the art of computerized chemistry, scientific knowledge is seamlessly translated into digital form without any need for human interference. Dealing with chemistry conundrums inspires us to explore innovative AI techniques that can be effectively utilized to tackle complex chemical issues requiring specialized knowledge. An innovative program utilizes chemical expertise to address challenges in the field of chemistry, specifically focusing on structure, design, properties and synthesis. Utilizing expert systems, artificial neural networks and machine learning, our computer systems are equipped to handle vast amounts of chemical information. This review aims to explore the application of AI in reaction prediction and chemical synthesis.

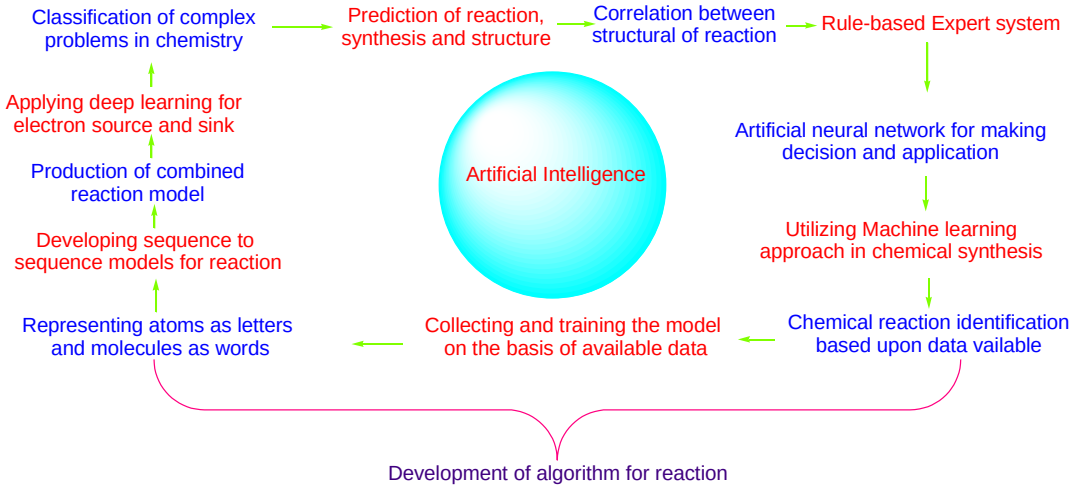
Keywords: Artificial intelligence, Organic reactions, Artificial neural network, Machine learning, Expert system, Complex reaction prediction.

Introduction

The ability to think and understand automaticity and act accordingly is intelligence. The goal of incorporating AI into chemistry is to create machines with human-like intelligence[13,24,40]. Artificial intelligence is the science enabling computers and machines to learn, reason and act in such a way that generally requires human intelligence. AI focuses on artificial device intelligence for analyzing data on a broad scale. McCarthy introduced AI in 1956. Like human intelligence, a machine that acts like a human is smarter than a human as it can handle broad and complex data without getting tired[3,28,36]. As in chemistry, machines may have different levels of intelligence. Computer intelligence excels at cognitive tasks like humans.

Human experts in a field must be consulted to create a truly intelligent computer system. This system should help humans to find information, make decisions, solve complex problems and understand sentences[12,16]. Figure 1a represents the main approaches to artificial intelligence. Artificial intelligence solves problems with algorithms. Physical and life scientists are rapidly adopting these methods. Without logic and problems, science is impossible.

Scientists care more about solution quality than method when solving a new problem. The goal is to create methods where computers intelligently understand the language, methodology, facts and logic behind input using massive amounts of data[7,39,54,57,83]. Figure 1b shows few of the most notable artificial intelligence applications. Software that has human-like behavior like conversation and language understanding is artificial intelligence. Experimental scientists want problem-solving software. They prefer AI for problem-solving[14,32]. Artificial neural networks (ANNs) and genetic algorithms are learning-based methods that are transforming science. Artificial intelligence depends on learning, which can be explained by process rule-based expert systems, where a science expert feeds the system the most relevant information over time[55,79].



Graphical abstract

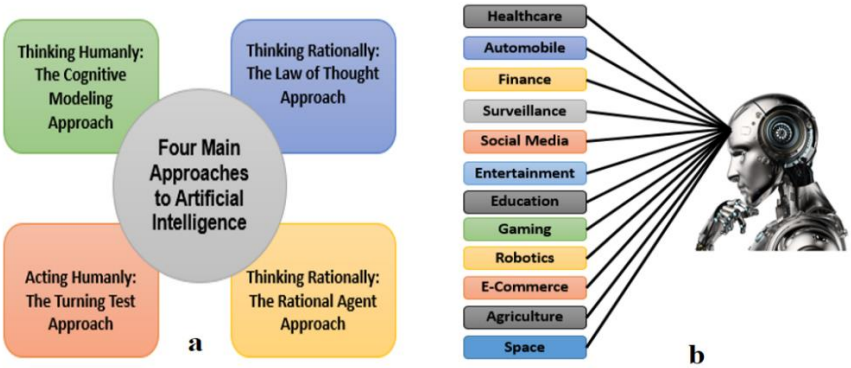


Figure 1: (a) Methodologies that are utilized in artificial intelligence, (b) Applications of AI in various fields

Artificial Neural Network: They examine a great deal of examples derived from logic and rules, which is necessary for comprehension and learning through the extraction of examples from databases.
Classifier System: They receive feedback from it and learn from every mistake.
Self-Organizing Map: Makes deduction just by looking at data without knowing what is expected to learn.

Application of AI in chemistry
Chemical synthesis is one of the most prominent fields contributing to environmental pollution and needs to be revolutionized for sustainable development. We have been continuously working in development of cleaner, greener and easier protocol for the synthesis of biologically potent scaffolds[61-69]. For revolutionizing the chemical synthesis and developing sustainability in the chemical synthesis, we look upon the Artificial Intelligence, the probable solution. But it is a very hard task and impractical to go through every problem like synthesis, separation, purification, conformational analysis, examining bioactivities in chemistry. So, for scientists, some search methods are required what make intelligent decisions about what to accept and which to ignore. Finding effective medications from the vast array of compounds that could have been synthesized is excellent evidence for the application of AI. AI is useful in the following contexts:

Classification of complex problems: Our job is to organize complex data like medical test results and chromatography analysis results. Interpreting the data can be difficult because it may be complex and produce conflicting or inconclusive results during rigorous testing[47,60].

Prediction of reaction, synthesis and structure: The identification of stable species that can be formed from many atoms is computationally very costly because it requires much time, but still these studies are so important because they are used to understand the properties of small structure in nano chemistry[50,80].

Correlation: Future research will examine how organic contaminants like poly chlorinated biphenyls' structural makeup affects their biodegradability. The empirical evidence supporting this relationship is still incomplete. Self-organizing maps can predict PCB degradation and stability without experimental data. You can better understand biodegradation. This method studies lipid bilayer properties for molecular dynamics simulations to understand their interactions in natural systems and biosensors[44,78]. Identification of chemical reactions is a widespread issue. Major reaction identification methods include:

Rule Based Expert System: How is knowledge defined? A subject or domain is understood theoretically or practically. People with extensive knowledge are called experts. Domain experts have deep knowledge and practical understanding in a field[4,11]. Expert systems are one of the most successful AI commercial applications. It solves specific problems using domain-specific knowledge to perform expertly in an application area. Narrower problems have higher success rates[4,11]. This type of personal consultant engages users in discussions to provide expert advice. Expert systems interact with users in such a way that users almost do not realize that they are talking to a computer. An intelligent expert system converses with humans[1]. An IF-THEN rule, which provides information in the IF part and an action in the THEN part, can be characterized as knowledge[1,17].

They pay attention to a selected specialist topic and know nothing outside that. They create this obscure view of life by having a subject knowledge that few people can quantify and complete with a thinking skill that they act as a human expert, permitting deduction of data provided by the user accordingly. These are based on heuristic knowledge[22,74].

The expert system extracts information from human specialists and processes it to solve problems like an expert. Figure 2 compares human and expert system. Information like objects, facts, data and rules can be used to manage and provide computer program-friendly information. Technological solutions are difficult without a thorough understanding of the problem and how to solve it. The purpose of these systems is to give suggestion, information and solve problem. They act like an expert, in some cases beyond expert performance. It should be able to reason the facts and rules in the form of symbols. It is based on rules and facts extracted from expert knowledge. Expert systems must defend their decisions like humans. The medical expert can answer our questions and can explain why. Like human

experts, expert systems are expected to provide some explanation, at least close to it[22,74]. The elements used in ES are shown in fig. 3.

Rule interpreters read input data, find applicable rules and apply them. Applying rules generates case-specific data. By developing expert systems, we can expect them to solve problems faster than humans. These systems use database rules, facts and relationships to make deductions and suggest based on user input[22,74]. Figure 4 shows an expert system and medical expert talking like a human expert and non-specialist. Conversation is happening as two people talk. It is not enough for the expert system to act like a human and follow human deductive pathways. The machine does not know the user, but it can identify the problem based on user input. It can deduce user-relevant information from input using its knowledge, rules and facts[22,74]. Rule-based feedback prediction can be fast, but developing and maintaining complex rule systems takes time.

Although law-based systems can work for some chemicals, their effectiveness is limited by human experts' rules. Note that there is no comprehensive chemical law system[2,35,81]. Performance of rule-based systems gets deteriorated in the long run when there is addition of new chemical properties. Additionally, these systems predict general chemical conversion. Multistep responses are summarized in one version, but arrow-clicking actions, including multiple-response responses, are not. These basic steps are building blocks to predict global responses to multiple measures and identify products[2,35,81].

Artificial Neural Network

Most chemical issues involve complex systems. Understanding and predicting chemical reactions is essential to organic chemistry and operational strategy[52]. Many chemical processes can be explained mathematically. Due to rapid technological advancement, today's world uses a wide range of advanced software and algorithms for commercial and educational purposes[45]. Most machine learning methods allow computers to learn through experience, exemplification and analogy. Intelligent systems can improve performance through continuous learning by understanding growth potential. Like chemists, mechanical learning methods underpin adaptable programs. Artificial neural networks and genetic algorithms are widely recognized as the most popular methods for machine learning[15].

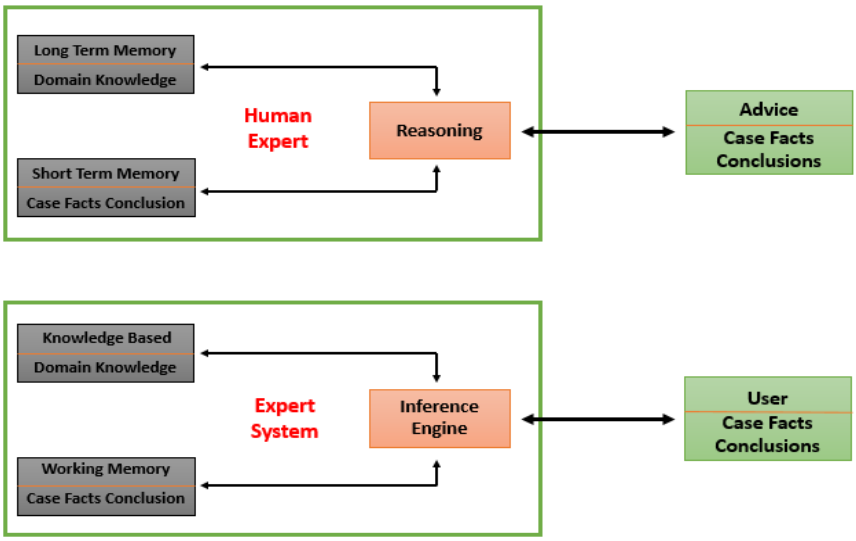


Figure 2: Comparison between human expert and expert system

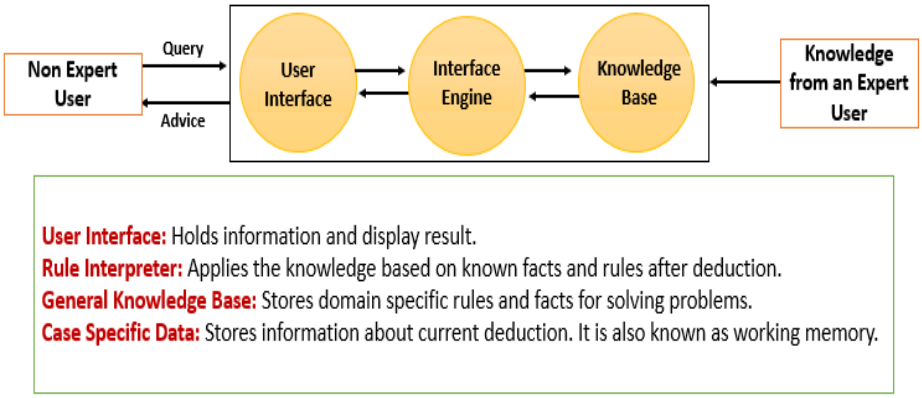


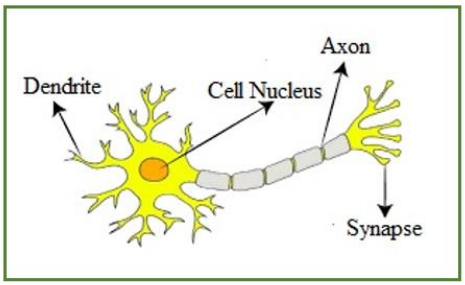
Figure 3: Architectural design for expert systems

In addition to fast computational speed, artificial neural networks can read and process data, tolerate errors and create data naturally. 90s studies show that ANN outperforms other mathematical methods[15]. ANN means neurons connected to a network which performs sharing of data. It is a system that has been inspired biologically. Human brains have many neurons that form a complex 3D network. Human nervous system has 1010 neurons. Nuclei make up neuron cell bodies. Further cell body divides into two parts – dendrites and axon. Large-surfaced dendrites receive and send signals to the cell body. Collaterals on the cell body's axons send signals to other neurons. Synapses connect axons and collaterals to dendrites or cell bodies of other neurons[26,59]. The dendrites and axon pass the signal electrically while the synapse passes the signal through chemical substances (Fig. 4).

Dendrites in ANN resemble user input signals, like in biological networks. After receiving user input, ANN transfers data to the neuron for computation and training, just like nervous system dendrites send signals to the cell body (soma). The output from the first may be the input for the other neural network[3,18]. Activation functions used in ANN for better output are shown in figure 5. These functions are step function, sign function, sigmoid function, linear function.

ANNs learn from data instead of human experience like human brains do. As the brain learns from experiences and trains data based on performance, learning algorithms work. We want machines to copy the human brain's process and function. Artificial neurons are connected like a network to learn and train data for better output. The network of neurons in humans helps to send information and data to different parts of the body quickly. ANN can recognize handwritten patterns that experts cannot. ANN has many interconnected processors called neurons. These neurons are signals between neurons; they receive many inputs but output one[27,53,72]. The artificial neural network aims to give machines brain-like networks that learn and act based on stored data (Fig. 6).

Input Layer: receives user input and sends it to neurons in the hidden layer.
Hidden Layer: receives the signals from input layer and learning, training algorithm is being done by extracting the information on the basis of data training.
Output Layer: receives the signals from hidden layer and project to the user.



Comparison between BNN and ANN	
Dendrites	Input signals
Cell body	Neurons
Synapse	Weighted links
Axon	Output signals

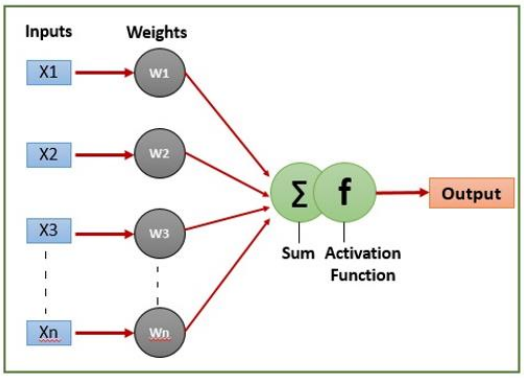


Figure 4: Resemblance of biological neural network and artificial neural network

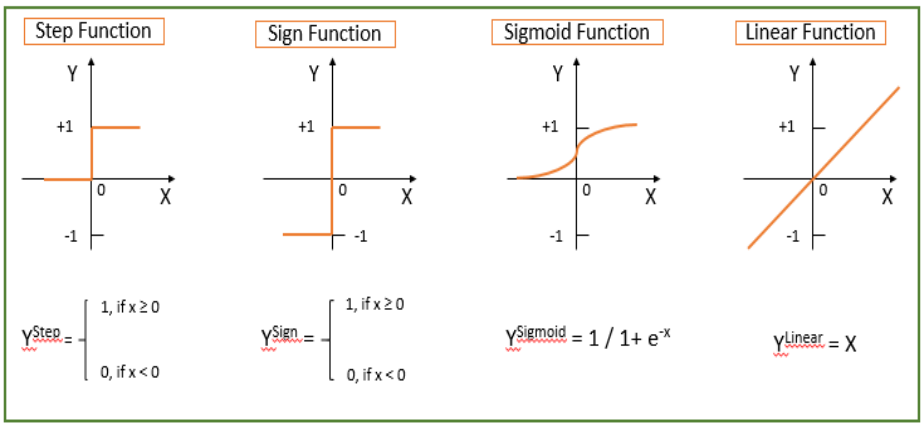


Figure 5: Activation functions for ANN

ANN construction requires knowing the number and connections of neurons. After that, choose an algorithm. The neural network training process concludes. The human brain's complexity has always fascinated scientists, who want to understand it. The brain neural network has information processing units that resemble neurons, like a chemist[6,76]. The fabricated neuron excels at basic mathematical operations, but its true power lies in its ability to form a network with other neurons. These systems are used for object classification, functional relationship modeling, data storage and retrieval and large data representation. They excel at chemical data processing, spectroscopic analysis, reaction prediction, chemical process control and electrostatic energy analysis[33,38,70,75].

The neural network model was inspired by how human brains process information through connected networks. Supercomputers cannot match the brain's processing power. Modern computers follow a program and algorithm sequentially, while human brains process information concurrently. Like a lab scientist, the human brain can recognize a friend's face at a glance or react to danger in a split second. ANN software for Von Neumann computers is available. The correlation between the infrared spectrum and chemical structure could be investigated using the same algorithm[27,53,72]. A network of basic processing components connected by weighted links is shown in figure 8. The computation unit computes the inputs after the input units process them.

Neural networks are capable of solving problems of this nature.

Classification: Assigning object identification to specific categories based on different attributes.

Modeling: A neural network can produce binary and real values. Integrating experiment results can yield new insights and advance the field. Using precise calculations, mathematics establishes a relationship.

Association: Neural networks are great for associative tasks like comparing information between related objects. Their ability to store similar data allows this.

Mapping: Here more complex information is converted into simpler representation[46].

Machine Learning Approaches

Although fast and helpful in identifying reactions, it requires a lot of reaction knowledge to learn. As such data is not available for academic use, collecting such a large amount of chemical reaction data is difficult. Today's deep learning and machine learning make reaction identification easier[42]. It projects chemical reactions using deep learning. Chemical reactions are projected at their individual steps in a complex reaction and adding each elementary step creates a complex reaction. Once we have a starting material and target mass, the system searches for unknown products (design, production) and suggests structures[20]. We can tell how projection is proposed and how the step by step reactions are involved. Reaction prediction collets all the precisely developed data for practicing and work in the field of modifying set of data as much as available.

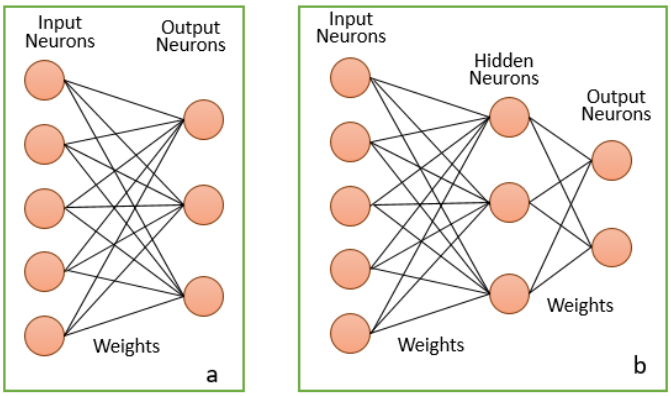


Figure 6: (a) Single layer perceptron with one hidden layer, (b) Working architecture of artificial neural network

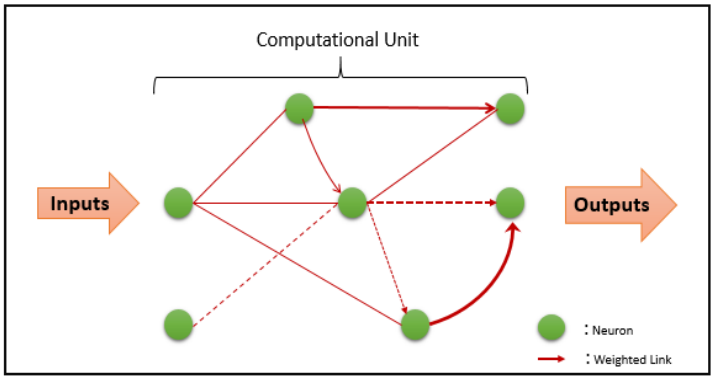


Figure 7: Neural network of simple processing element

Reaction prediction predicts chemical reactions and their paths. Deep learning is used to study fundamental reactions. Combining all basic reactions helps explain complex reactions. Rule-based systems may provide machine learning datasets[20,42]. By utilizing the findings from the QM approach, one can enhance the ML system by incorporating them into ML algorithms, thus expanding the base knowledge. Understanding reactions is a complex process that involves deep learning at every step. The method demonstrates how humans estimate chemical reactions[56]. Each step involves electron movement from source to sink. Complex reaction identification involves following 4 steps:

- Identifying all the practicable electron source and sinks in loading the reactant species.
- Separating all the sources and sinks taking out some which are highly reactive.
- Showing all the feasible compounds of source-sink pairing.
- Naming the suggested reaction according to their convenient conditions.
- Repeat the following processes to predict the complex reactions for finding unknown products.

Chemical Reaction Identification

Reaction identifier can be discussed on the basis of reference data on how they perform, on emerging actual world reaction and exhibit a high degree of perfection. Hochreiter and Schmidhuber[29] correlated the performance of prior samples and found a new way to identify electron source and sink using LSTM (Long Short Term Memory), with favorable results. Chemistry explores infinite chemical space. If we like multiple complex reactions, artificial intelligence (AI) uses neural networks to advance organic synthetic chemistry[34].

Schwaller et al[58] discussed forward chemical reaction identification problems using neural sequence-to-sequence model. Given a set of reactants, reagents and conditions, the problem is identifying the most likely product.

The reaction mentioned in figure 8 is one part of the interesting problem in chemical industry. When one started identification of forward reaction problem, one of the major thought is that we are strong believer that in chemistry, there is specific language. The grammar of these languages may be complex or difficult for the human brain to define precisely[58,82].

Collection and training of data: Development of AI involved treating chemistry like a language. Many AI tools translate these languages, but Schwaller et al[58] tried to develop a similar tool for chemistry, translating reactants, reagents and conditions into possible products[23]. Data overload was the first problem. AI requires data to train the model and chemistry requires data to train reactions and develop ideas.

Here, we are representing the excellent work in the field of natural language processing conducted by Lowe et al[43] at the University of Cambridge (Fig. 9). Textual information from reaction patents is extracted in this work. The reaction described in patent procedure was intrinsic text-mining activity to the generation of SMARTS databases[30]. Lowe et al[43] started from these databases and performed a series of chain operations to a certain level of quality in the initial training of data to create a dataset of approximately millions of chemical reactions, which they and other groups use to compare the quality and validity of different AI models for identifying chemical reactions[54,73].

Representation: Atom as Letters and Molecule as Words

Despite being far from the molecular graph and SMILES string center, functional groups, solvents and catalysts can significantly affect a reaction's outcome. Figure 10 shows the network's ability to prioritize the C[O$^-$] molecule, accurately match the input's [O$^-$] to the target's O and ignore the target's Br substitution[58].

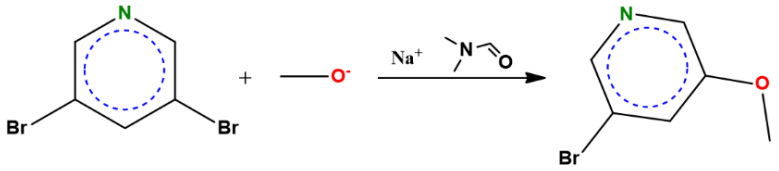


Figure 8: Identification of product with the given input

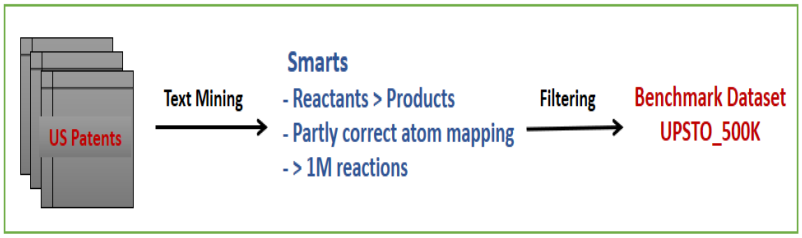


Figure 9: Collection and training of data

Working Sequence-2-Sequence Model

Figure 11 shows that in the left part of the reaction, reactant, reagent and conditions are encoded as a specific encoder into features that are decoded as generative models to produce chemical reaction outcomes. All encoder-features-decoder layers must be tuned during chemical reaction training. Fig. 11 shows training, which tunes all features. For accurate intermediate AI model training, Schwaller et al[58] used millions of data sets. Since input is known, output is too. Thus, by observing input and output, one can determine which output matches which input and optimize the chemical reaction intermediate[15,43].

Reaction predictor is an advanced method for predicting fundamental chemical reactions. Fooshee et al[21] described the ML architecture and process for reaction predictor ML predictions[21,71]. They carefully created a training database with over 11,000 fundamental reactions from novel organic chemistry. The researchers also present a promising LSTM-based reactive site prediction method using only SMILES strings and compare it to a prototype method.

Collection of Data: The data set contains simple reactions. Every elementary reaction has an energetically favorable mechanistic step with one transition state. It has one labeled electron sink and source. Chemical reactions can be predicted using polar, radical, or pericyclic methods. These methods identify chemical reactions differently[10]. To evaluate the system's effectiveness, Fooshee et al[21] used 289 fundamental reactions from the 5551 elementary reactions in the main data. Since polar data and models predict reactions, they examined them. From the literature response and strategic use of organic synthesis reaction, the reactions were selected to cover a wide range of high-quality biological and chemical substances to test the program's ability to add to the actual response[41].

Development of Data Set: Reaction Explorer rules have a major drawback: they require many training results to simulate undergraduate chemical exposure bias. Training covered only first, second and third lines. Undergraduate texts simplify or leave complex details to explain fundamentals. Research is needed to determine chemical truth. Learning algorithms that use training data have this issue. Fooshee et al[21] fully set the data. After manually testing all principal responses, 884 problem responses were removed. Reaction has faded for various reasons. Basic response duplication reached 10%. Second, incorrect arrow press formatting removed reactions. Third, training reactions had essential products. The questionable response was removed, leaving 4,667 primary polar response data.

From this refined data, they added 6,361 top-rated hand-held reactions[41]. They added over two thousand reactions to basic training to fill gaps in active phosphorus, sulphur and active groups, silicon, proton transfer and tetrahedral intermediate. About 1,500 more low-level responses were selected from response methods and first-year biological studies on non-classic carbocations, refined rings, vinylsilanes, sulphur, carbonyl, phosphorus, compound, enols, enamine, allylsilanes and more. The well-known answer book yielded 1,004 technical measures. Nearly 2000 chemical and research presentation outcomes were randomly selected to complete the data set.

Figure 12 illustrates some of the data set's remarkable new responses. Figure 12a depicts eight-step solid reactions. Seven mechanical steps comprise the Mitsunobu reaction in figure 12b. Figure 12c depicts 12 mechanical MPV reduction steps. These critical answering machine steps require a basic response that should not have been in the data set. Basic data set responses were 11,028. The system improves with data[71].

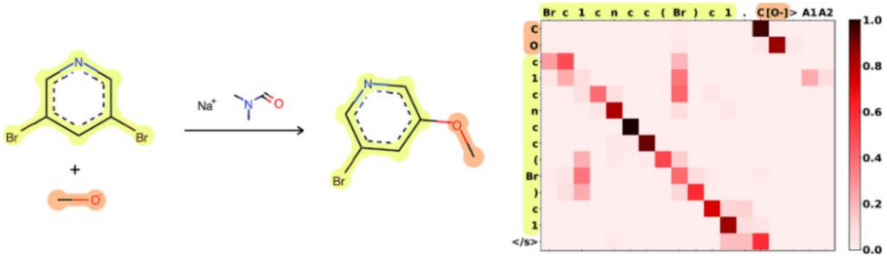


Figure 10: SMILES TO SMILES prediction with SEQUENCE-2-SEQUENCE model[58]

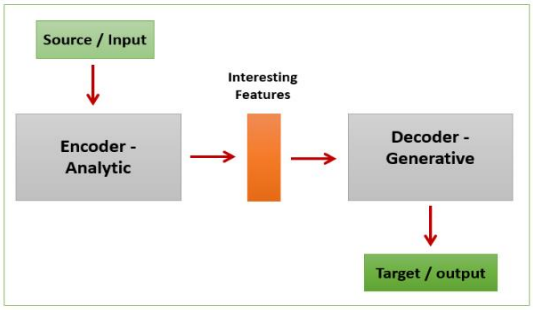


Figure 11: Architecture of SEQUENCE-2-SEQUENCE of model

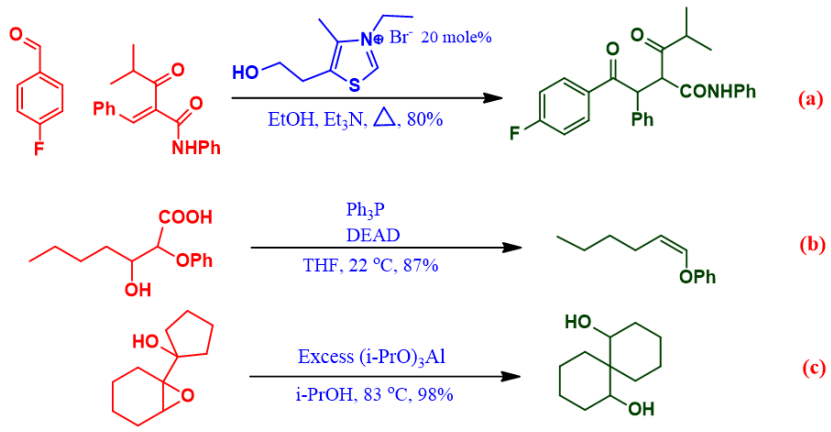


Figure 12: (a) An eight step Stetter reaction[8] (b) A seven-step Mitsunobu reaction[48] (c) A 12-step MPV reduction[77]

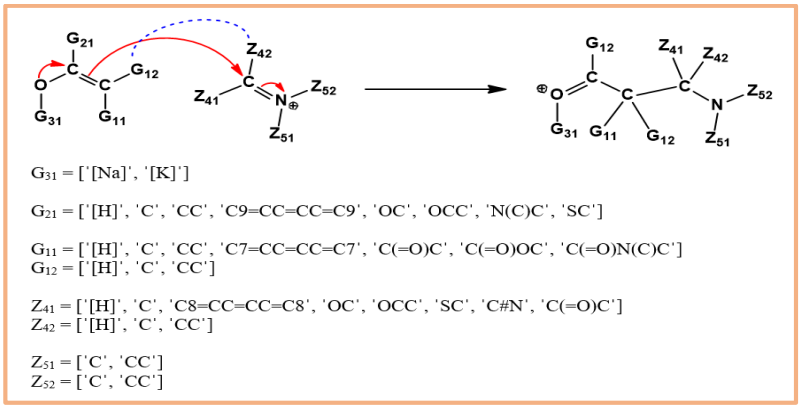


Figure 13: Visual representation of the process of generating combinatorial reactions, featuring substitution constraints at the bottom and a reaction template at the top. Countless reactions can be generated by exploring all possible substitutions at the G and Z** sites**

Production of Combined Reaction: Fooshee et al[21] attempted to utilize programs that autonomously generated a multitude of fundamental responses to the training data by employing the subsequent approach. First, through the given response process, they determined the primary molecular templates and the appropriate electron migration. The compound is formed through a systematic response of the template throughout the chemical realities, resulting in the complete composite compound. They created a comprehensive catalog of essential response categories using their expertise. By sampling 1,000 reactions per mechanism at random, they avoided biasing data toward combined reactions. Figure 13 depicts the entire process.

Applying Deep Learning: Understanding the involved electron sinks and sources is a crucial part of the Response Predictor pipeline. If reliable sink or a high-quality source is disregarded during the sink/source sorting process, it becomes impossible to replicate the desired response. Reaction predictor initially prioritized memorization over accuracy. Even though most of the predicted sinks and sources were false, the system preferred to cover all bases. However, false positives can harm operations. They can greatly increase the time it takes to perform calculations, especially when it comes to method search. This is because single-step predictions need to be connected in order to predict multi-step products. In their study, Fooshee et al[21] constructed source/submerged filtering models that prioritize accuracy and memory, as highlighted in figure 14. Through various experimentation methods, they have successfully achieved outstanding results using a single completely linked feed-forward neural network. This network consists of 1500 input data and hidden layers containing 200 units with two separate sigmoid output units that correspond to a prediction for a sink and source. A 50% exit was applied to each hidden layer[5] and an early stop was used. The Adam optimizer was used to reconstruct 64 models in small batches to make the model work following Glorot and Bengio[25].

Based on 10% training set verification, rotten and premature learning was employed. Cameras and Tensor Flow power the models and NVIDIA Titan X GPUs train them. This is how they built the training network data source/sink filter. They included four atomic reconstructing training examples for all primary database responses: two random no-sources, no-sink illustrations, a labelled sink and a labelled source. They specifically steer clear of instances with poor labels that are better suited as second sources or sinks. Considering the basic response, a set of atoms that are not specifically mentioned in a sink or source, comprises a lot of inefficient sinks and resources.

However, it may be a source or sink of the "second phase" when viewed alone or in specific cells. A model representing a non-source instance using data can be created by randomizing the atom selecting process in negative constructions. This prevents us from labeling all second-class sources and sinks as negative models. According to these numbers, the training kit's filter removed 23,850 samples, half of which were good examples. They listed the outcomes of the recommended suggestion after determining the sinks and sources and connecting them. In order to compute the points used, they train an extensive network of Siamese architecture[19,31]. The structure is depicted in figure 15.

The training examples have two sets of sorted responses (Rfavorable, Runfavorable), with the randomized weight network instance always useful. Fixed weights are +1 and -1 to the left and right of the results for the sigmoid unit. Thus, if the left response scores higher than the right answer, y is closer to 1 and if lower, 0 is obtained. They use a single average weight network model to calculate target scores for each response after modeling training and evaluate them. They combined two hidden layers of 300 tanh units with the gmoid extraction in a distributed weight of the network. As mentioned, the source/sink model will be implemented and trained. Next, they finished Rfavorable and Runfavorable training examples.

Every primary response to a set of data with matched electron sources and labelled electron sinks receives a favorite response. In terms of chemical feasibility, they caused many negative reactions by matching all non-source sources and labelled sinks with well- and poorly-labeled sources. They use this negative feedback set to generate 387,744 training instances for pairs (Rfavorable, Runfavorable).

Feature Representation and Selection

The pertinent chemical information about every possible source or sink potential inside the active molecule must be extracted in order to create a precise source or sink. Fooshee et al[21] employed two types of features, physico-chemical and graph-topological, to collect this data. Atomic level extraction is used to extract physicochemical properties. A steric coefficient, lone pairs, filled and unfilled orbitals and partial and formal charge are a few examples. Graph-topological attributes describe the atom's characteristics and the molecular graph's bond connectivity. Based on numerous chemical fingerprints, it was published when new drugs and technologies were introduced in a limited area surrounding a specific atom[54].

In theory, they had performed depth-first methods up to six atoms if those atoms are heteroatoms or part of an extended pi system; if not, the optimum size is three. Next, they extracted the response rate in order to identify the general cell (s) that are involved in the principal response as well as the source and sink of the spontaneous response. Only then can we accurately measure spontaneous response. These response attributes interact a combined source with atomic-level properties, features defining the orbitals and fingerprints to accept reaction changes including those involving active groups and subjects being created and destroyed.

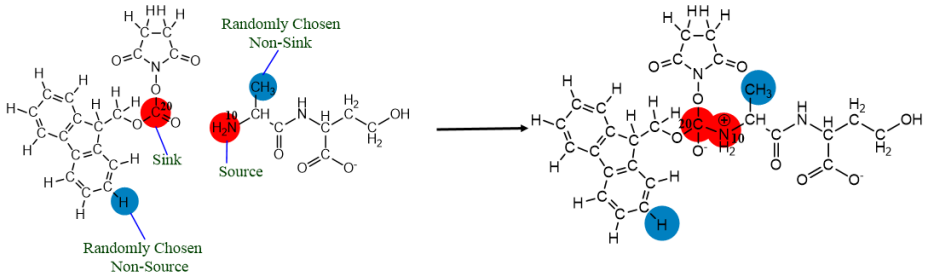


Figure 14: Atom reactivity training examples retrieved for the sink/source sorting network

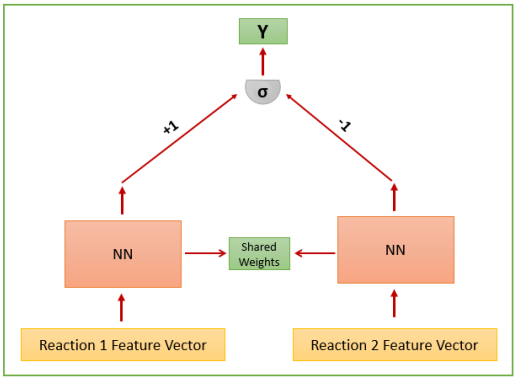


Figure 15: Siamese representation for ranking of reaction

Subtracting the above symptoms from all data instances yields 293,046 atomic features and 62,560 reaction levels. They select the top 2,000 reaction components and 1,500 atomic levels from the best data and submit them to sink/source and assessment models.

The primary goal of incorporating artificial intelligence into chemistry is to create machines that can operate with the same level of intelligence as humans. Computer intelligence is its ability to perform cognitive tasks like humans. We need human experts in a field to build a truly advanced system. This system should aid in information retrieval, decision-making, problem-solving and sentence comprehension. So, individuals can effectively address their scientific challenges[37,51].

Conclusion

Advancements in AI within the realm of chemistry continue to be a highly complex subject. Further advancements and research are necessary in the realm of AI to enhance its effectiveness and simplify its application in solving chemical problems. Additionally, it is crucial to invest in the development of new research tools and programs that can optimize the work of scientists and streamline their processes, ultimately saving valuable time. Using the emerging field of machine learning, we aim to analyze the impact of C–N cross-linking and its effectiveness in chemical engineering. Our focus is on illustrating the advancements of artificial intelligence in this domain. Mastering molecular synthesis remains a crucial endeavor in the realm of organic chemistry. It requires a composer's strategic approach to problem-solving, drawing from their experience.

However, the process itself can be repetitive and time-consuming, often resulting in suboptimal solutions. Given the recent remarkable advancements in machine learning, it is clear that investing in organic chemistry will greatly expedite the discovery of new drugs and will propel the future of the basic sciences. This study explores various methods that shed light on the significant role computers play in the field of chemistry, aiding chemists in their work.

References

1. Ahmad S.F., Han H. and Alam M.M., Impact of artificial intelligence on human loss in decision making, laziness and safety in education, *Humanit Soc Sci Commun*, **10**, 311 **(2023)**

2. Aldoseri A., Al-Khalifa K.N. and Hamouda A.M., Re-Thinking Data Strategy and Integration for Artificial Intelligence: Concepts, Opportunities and Challenges, *App Sci*, **13**, 7082 **(2023)**

3. Anand A., Sen S. and Roy K., Quantifying the Brain Predictivity of Artificial Neural Networks with Nonlinear Response Mapping, *Front Comput Neurosci*, **15**, 609721 **(2021)**

4. 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**, 2357 **(2020)**

5. Baldi P. and Sadowski P., The Dropout Learning Algorithm, *Artif Intell*, **78**, 78-122 **(2014)**

6. Basheer I.A. and Hajmeer M., Artificial neural networks: fundamentals, computing, design and application, *J Microbiol Methods*, **43**, 3-31 **(2000)**

7. Baum Z.J., Yu X., Ayala P.Y., Zhao Y., Watkins S.P. and Zhou Q., Artificial Intelligence in Chemistry: Current Trends and Future Directions, *J Chem Inf Model*, **61**, 3197 **(2021)**

8. Baumann K., Butler D., Deering C.F., Mennen K.E., Millar A., Nanninga T.N., Palmer C. and Roth B., The convergent synthesis of CI-981., an optically active., highly potent., tissue selective inhibitor of HMG-CoA reductase, *Tetra Let*, **33**, 2283-2284 **(1992)**

9. Benvenuti M. et al, Artificial intelligence and human behavioral development: A perspective on new skills and competences acquisition for the educational context, *Comput Hum Behav*, **148**, 107903 **(2023)**

10. Carey F.A. and Sundberg R.J., Advanced organic chemistry, part a: Structure and mechanism, Springer, New York **(2007)**

11. Chen J.H. and Baldi P., No electron left behind: a rule-based expert system to predict chemical reactions and reaction mechanisms, *J Chem Inf Model*, **49**, 2034-2043 **(2009)**

12. Chiu T.K.F., Xia Q., Zhou X., Chai C.S. and Cheng M., Systematic literature review on opportunities, challenges and future research recommendations of artificial intelligence in education, Computers and Education, *Artif Intell*, **4**, 100118 **(2023)**

13. Choudhary N., Bharti R. and Sharma R., Role of artificial intelligence in chemistry, *Mater Today Proc*, **48**, 1527-1533 **(2022)**

14. Chubb J., Cowling P. and Reed D., Speeding up to keep up: exploring the use of AI in the research process, *Artif Intell Soc*, **37**, 1439–1457 **(2022)**

15. Coley C.W., Green W.H. and Jensen K.F., Machine Learning in Computer-Aided Synthesis Planning, *Acc Chem Res*, **51**, 1281-1289 **(2018)**

16. Collins C., Dennehy D., Conboy K. and Mikalef P., Artificial intelligence in information systems research: A systematic literature review and research agenda, *Int J Inform Manag*, **60**, 102383 **(2021)**

17. Da Silveira T.B.N. and Lopes H.S., Intelligence across humans and machines: a joint perspective, *Front Psychol*, **14**, 1209761 **(2023)**

18. Deng L., Wu Y., Hu X., Liang L., Ding Y., Li G., Zhao G., Li P. and Xie Y., Rethinking the performance comparison between SNNS and ANNS, *Neural Netw*, **121**, 294-307 **(2020)**

19. Dincă Lăzărescu A.M., Moldovanu S. and Moraru L., A Fingerprint Matching Algorithm Using the Combination of Edge Features and Convolution Neural Networks, *Inventions*, **7**, 39 **(2022)**

20. Elton D.C., Boukouvalas Z., Fuge M.D. and Chung P.W., Deep learning for molecular design-a review of the state of the art, *Mol Syst Des Eng*, **4**, 828-849 **(2019)**

21. Fooshee D., Mood A., Gutman E., Tavakoli M., Urban G., Liu F., Huynh N., Van Vrankenb D. and Baldi P, Deep learning for chemical reaction prediction, *Mol Syst Des Eng*, **3**, 442-452 **(2018)**

22. García-Peñalvo F.J., Vázquez-Ingelmo A. and García-Holgado A., Explainable Rules and Heuristics in AI Algorithm Recommendation Approaches—A Systematic Literature Review and Mapping Study, *Comput Mod Eng Sci*, **136**, 1023–1051 **(2023)**

23. Gasteiger J. and Jochum C., EROS A computer program for generating sequences of reactions, In Organic Compunds, Topics in Current Chemistry, Springer, Berlin, Heidelberg, 74 **(1978)**

24. Gasteiger J., Chemistry in Times of Artificial Intelligence, *Chemphyschem*, **16**, 2233-2242 **(2020)**

25. Glorot X. and Bengio Y., Understanding the Difficulty of Training Deep Feedforward Neural Networks. Proceedings of the Thirteenth International Conference on Artificial Intelligence and Statistics, *Proc Mach Learn Res*, **9**, 249-256 **(2010)**

26. Goel A., Goel A K. and Kumar A., The role of artificial neural network and machine learning in utilizing spatial information, *Spat Inf Res*, **31**, 275–285 **(2023)**

27. Guest O. and Martin A.E., On Logical Inference over Brains, Behaviour and Artificial Neural Networks, *Comput Brain Behav*, **6**, 213–227 **(2023)**

28. Hassani H., Silva E.S., Unger S., TajMazinani M. and Mac Feely S., Artificial Intelligence (AI) or Intelligence Augmentation (IA): What Is the Future?, *Artif Intell*, **1(2)**, 143-155 **(2020)**

29. Hochreiter S. and Schmidhuber J., Long Short-Term Memory, *Neural Comput*, **9**, 1735 **(1997)**

30. Holliday G.L., Murray-Rust P. and Rzepa H.S., Chemical markup, XML and the world wide web. 6. CMLReact., an XML vocabulary for chemical reactions, *J Chem Inf Model*, **46**, 145-157 **(2006)**

31. Hsiao C.T., Lin C.Y., Wang P.S. and Wu Y.T., Application of Convolutional Neural Network for Fingerprint-Based Prediction of Gender, Finger Position and Height, *Entropy (Basel)*, **24**, 475 **(2022)**

32. Joksimovic S. et al, Opportunities of artificial intelligence for supporting complex problem-solving: Findings from a scoping review, *Comput Edu.: Artif Intell*, **4**, 100138 **(2023)**

33. Käser S., Vazquez-Salazar L.I., Meuwly M. and Töpfer K., Neural network potentials for chemistry: concepts, applications and prospects, *Digit Discov*, **2**, 28-58 **(2022)**

34. Kayala M.A. and Baldi P., Reaction Predictor: prediction of complex chemical reactions at the mechanistic level using machine learning, *J Chem Inf Model*, **52**, 2526-40 **(2012)**

35. Kierner S., Kucharski J. and Kierner Z., Taxonomy of hybrid architectures involving rule-based reasoning and machine learning in clinical decision systems: A scoping review, *J Biomed Inform*, **144**, 104428 **(2023)**

36. Korteling J.E.H., Van de Boer-Visschedijk G.C., Blankendaal R.A.M., Boonekamp R.C. and Eikelboom A.R., Human- versus Artificial Intelligence, *Front Artif Intell*, **4**, 622364 **(2021)**

37. Kraskov A., Stögbauer H. and Grassberger P., Estimating mutual information, *Phy Rev E*, **69**, 066138 **(2004)**

38. Kulichenko M., Smith J.S., Nebgen B., Li Y.W., Fedik N., Boldyrev A.I., Lubbers N., Barros K. and Tretiak S., The Rise of Neural Networks for Materials and Chemical Dynamics, *J Phys Chem Lett*, **12**, 6227-6243 **(2021)**

39. Kulikowski C.A., Beginnings of Artificial Intelligence in Medicine (AIM): Computational Artifice Assisting Scientific Inquiry and Clinical Art - with Reflections on Present AIM Challenges, *Yearb Med Inform*, **28**, 249-256 **(2019)**

40. Kuntz D. and Wilson A., Machine learning artificial intelligence and chemistry: How smart algorithms are reshaping simulation and the laboratory, *Pure App Chem*, **94(8)**, 1019-1054 **(2022)**

41. Kurti L. and Czakó B., Strategic applications of named reactions in organic synthesis, *J Am Chem Soc*, **128**, 1032 **(2006)**

42. Li B., Su S. and Zhu C., A deep learning framework for accurate reaction prediction and its application on high-throughput experimentation data, *J Cheminform*, **15**, 72 **(2023)**

43. Lowe D.M. and Sayle R.A., LeadMine: a grammar and dictionary driven approach to entity recognition, *J Cheminform*, **7**, S5 **(2015)**

44. Lupyan D., Mezei M., Logothetis D.E. and Osman R., A molecular dynamics investigation of lipid bilayer perturbation by PIP2, *Biophys J*, **98**, 240-247 **(2010)**

45. Maryasin B., Marquetand P. and Maulide N., Machine Learning for Organic Synthesis: Are Robots Replacing Chemists, *Angew Chem Int Ed Engl*, **57**, 6978-6980 **(2018)**

46. Molnár L., Keseru G.M., Papp A., Lorincz Z., Ambrus G. and Darvas F., A neural network based classification scheme for cytotoxicity predictions: Validation on 30.000 compounds, *Bioorg Med Chem Lett*, **16**, 1037-1049 **(2006)**

47. Mukhamediev R.I., Popova Y., Kuchin Y., Zaitseva E., Kalimoldayev A., Symagulov A., Levashenko V., Abdoldina F., Gopejenko V. and Yakunin K., Review of Artificial Intelligence and Machine Learning Technologies: Classification, Restrictions, Opportunities and Challenges, *Mathematics*, **10(15)**, 2552 **(2022)**

48. Mulzer J., Pointner A., Chucholowski A. and Brüntrup G., Threo-3- Hydroxycarboxylic acids as key intermediates in a highly stereoselective synthesis of (Z)-and (E)-olefins and enol ethers, *Chem Comm*, **2**, 52-54 **(1979)**

49. Navratil J., De Paola G., Kollias G., Nadukandi P., Codas A. and Ibanez-Llano C., An End-to-End Deep Sequential Surrogate Model for High Performance Reservoir Modeling: Enabling New Workflows, Paper presented at the SPE Annual Technical Conference and Exhibition, Virtual **(2020)**

50. Paul D. et al, Artificial intelligence in drug discovery and development, *Drug Discov Today*, **26**, 80-93 **(2021)**

51. Pedregosa F., Varoquaux G., Gramfort A., Michel V., Thirion B., Grisel O. and Duchesnay E., Scikit-learn: Machine learning in Python, *J Mach Learn Res*, **12**, 2825-2830 **(2011)**

52. Peiretti F. and Brunel J.M., Artificial Intelligence: The Future for Organic Chemistry, *ACS Omega*, **3**, 13263-13266 **(2018)**

53. Pham T.Q., Matsui T. and Chikazoe J., Evaluation of the Hierarchical Correspondence between the Human Brain and Artificial Neural Networks: A Review, *Biology (Basel)*, **12**, 1330 **(2023)**

54. Rodgers C.M., Ellingson S.R. and Chatterjee P., Open Data and transparency in artificial intelligence and machine learning: A new era of research, *F1000 Res*, **12**, 387 **(2023)**

55. Sabir Z., Ben Said S. and Al-Mdallal Q., An artificial neural network approach for the language learning model, *Sci Rep*, **13**, 22693 **(2023)**

56. Sadowski P., Fooshee D., Subrahmanya N.A. and Baldi P., Synergies Between Quantum Mechanics and Machine Learning in Reaction Prediction, *J Chem Inf Model*, **56**, 2125-2128 **(2016)**

57. Sarker I.H., Machine Learning: Algorithms, Real-World Applications and Research Directions, *SN Comput Sci*, **2**, 160 **(2021)**

58. Schwaller P., Gaudin T., Lányi D., Bekas C. and Laino T., Found in Translation": predicting outcomes of complex organic chemistry reactions using neural sequence-to-sequence models, *Chem Sci*, **9**, 6091 **(2018)**

59. Shao F. and Shen Z., How can artificial neural networks approximate the brain, *Front Psychol*, **13**, 970214 **(2023)**

60. Soori M., Arezoo B. and Dastres R., Artificial intelligence, machine learning and deep learning in advanced robotics, a review, *Cognitive Robotics*, **3**, 54-70 **(2023)**

61. Sowjanya Pulipati and Srinivasa Babu Puttagunta, In Silico Prediction of Biomarkers from Amaranthus tricolor (L) Targeted against HIV-1 Protease, *Res. J. Biotech.*, **18(1)**, 7–14, https://doi.org/10.25303/1801rjbt07014 **(2023)**

62. Srivastava N., Kishore R. and Chaturvedi D., Novel and efficient method for the synthesis of cyclic trithiocarbonates employing Cs_2CO_3 and CS_2 system, *Res. J. Chem. Environ.*, **25(12)**, 142-148, https://doi.org/10.25303/2512rjce142148 **(2021)**

63. Srivastava N., Saxena M. and Shukla, Novel Synthesis of 5-Oxo-2-thioxo-2,5-dihydro-3-thiophenecarboxylate Derivatives in Non-aqueous Medium, *Asian J. Chem.*, **31**, 176-180, https://doi.org/10.14233/ajchem.2019.21622 **(2019)**

64. Srivastava N., Maaz S.M., Bajpai S., Singh A. and Pandey G., Synthesis, characterization and in-silico studies of novel heterocyclic organotellurium dithiocarbamates, *Res. J. Chem. Environ.*, **25(5)**, 170–177 **(2021)**

65. Srivastava N., Rathore B.K., Vishnoi R.K. and Bajpai S., A Simple and Convenient Method for Synthesis of Artemisinin Dimer through Aliphatic Nitro Compounds, *Asian J. Chem.*, **33**, 49-52, https://doi.org/10.14233/ajchem.2021.22918 **(2020)**

66. Srivastava N., Chaturvedi D., Kishore R., Kaur M. and Kaur R., Triton-B/CS2 mediated novel synthesis of substituted thioureas: A novel class of anti-cancer agents, *Indian J. Chem.*, **61**, 870-877, DOI: 10.56042/ijc.v61i8.65053 **(2022)**

67. Srivastava N., Tetrabutylammonium Iodide/I_2Mediated Convenient and Green Synthesis of Substituted Organic Isothiocyanates, *Org Prep. Proced. Intern.*, **53**, 562-570, https://doi.org/10.1080/00304948.2021.1975487 **(2021)**

68. Srivastava N. and Saxena M., Novel Synthesis of Symmetrical Dialkyl/ Diaryl trithiocarbonates in Non-aqueous Medium at Room Temperature Using CS_2. CS_2CO_3 and Alkyl/aryl halide, *Rasayan J Chem.*, **12**, 333-337, http://dx.doi.org/10.31788/RJC.2019.1215038 **(2019)**

69. Srivastava N., Efficient TBAI-CS2 Promoted Synthesis of Substituted Hydrazinecarbodithiolates, *Org Prep Proced Intern.*, **54**, 569-577, DOI: 10.1080/00304948.2022.2111171 **(2022)**

70. Staub R., Gantzer P., Harabuchi Y., Maeda S. and Varnek A., Challenges for Kinetics Predictions via Neural Network Potentials: A Wilkinson's Catalyst Case, *Molecules*, **28**, 4477 **(2023)**

71. Stocker S., Csányi G. and Reuter K., Machine learning in chemical reaction space, *Nat Commun*, **11**, 5505 **(2020)**

72. Surianarayanan C., Lawrence J.J., Chelliah P.R., Prakash E. and Hewage C., Convergence of Artificial Intelligence and Neuroscience towards the Diagnosis of Neurological Disorders—A Scoping Review, *Sensors*, **26**, 3062 **(2023)**

73. Swamidass S.J., Chen J., Bruand J., Phung P., Ralaivola L. and Baldi P., Kernels for small molecules and the prediction of mutagenicity, toxicity and anti-cancer activity, *Bioinformatics*, **1**, i359-68 **(2005)**

74. Taherdoost H. and Madanchian M., Artificial Intelligence and Knowledge Management: Impacts, Benefits and Implementation, *Computers*, **12**, 72 **(2023)**

75. Taylor C.J., Pomberger A., Felton K.C., Grainger R., Barecka M., Chamberlain T.W., Bourne R.A., Johnson C.N. and Lapkin A.A., A Brief Introduction to Chemical Reaction Optimization, *Chem Rev*, **123**, 3089-3126 **(2023)**

76. Teshnehlab M. and Watanabe K., Fundamentals of Neural Networks, In Intelligent Control Based on Flexible Neural Networks, International Series on Microprocessor-Based and Intelligent Systems Engineering, Springer, Dordrecht **(1999)**

77. Tu Y.Q., Yang L.M. and Chen Y.Z., A Facial One-step Approach to Stereospecific Spirocyclic Diols from α-Hydroxyepoxides, *Chem Let*, **27**, 285–286 **(1998)**

78. Vattulainen I. and Rog T., Lipid simulations: a perspective on lipids in action, *Cold Spring Harb Perspect Biol*, **3**, a004655 **(2011)**

79. Velichko A., Korzun D. and Meigal A., Artificial Neural Networks for IoT-Enabled Smart Applications: Recent Trends, *Sensors*, **23**, 4853 **(4653)**

80. Venkatasubramanian V. and Mann V., Artificial intelligence in reaction prediction and chemical synthesis, *Curr Opin Chem Eng*, **36**, 100749 **(2022)**

81. Weiss S.M. and Indurkhya N., Rule-based Machine Learning Methods for Functional Prediction, *J Artif Intell Res*, **3**, 383-403 **(1995)**

82. Williamson E.M. and Brutchey R.L., Using Data-Driven Learning to Predict and Control the Outcomes of Inorganic Materials Synthesis, *Inorg Chem*, **62**, 16251-16262 **(2023)**

83. Xu Y., Liu X., Cao X., Huang C., Liu E., Qian S., Liu X. and Zhang J., Artificial intelligence: A powerful paradigm for scientific research, *Innovation (Camb)*, **2**, 100179 **(2021)**.

(Received 18th January 2025, accepted 22nd March 2025)

*****