Al-Imam Muhammad ibn Saud Islamic University
College of Science
Department of Chemistry

Machine Learning (ML) and Artificial Intelligence (AI) in organic Chemistry

A graduation research project
Submitted to the Department of Chemistry in partial fulfillment of the requirements
for the completion of the degree of Bachelor of
Science in Chemistry

BY
Basmah Abdullah Aldubaikhi
Basmah Yahya Abumaghayed
Abeer Obaid Alanazi

Under supervision of
Dr. Arwa Sultan Al Mubarak

Second semestr, February 2025

Table of Contents

Machine Learning (ML) and Artificial Intelligence (AI) in organic Chemistry

Abstract:

Machine learning (ML) and artificial intelligence (AI) are making deep changes, especially in the field of Chemistry, By collecting information more accurately and quickly. ML and AI are reshaping and accelerating scientific discovery and providing new insights. This review provides an overview of ML and AI in Organic Chemistry and focuses on ML and AI technologies, applications, and future trends in organic chemistry.

الخلاصة:

ان التعلم الالي والذكاء الاصطناعي يحدثان تغييرات عميقة، لا سيما في مجال الكيمياء، من خلال جمع المعلومات بدقة وسرعة اكبر. يعمل التعلم الالي والذكاء الاصطناعي على إعادة تشكيل وتسريع الاكتشاف العلمي وتوفير رؤى جديدة. تقدم هذه المراجعة نظرة عامة على التعلم الالي والذكاء الاصطناعي في الكيمياء العضوية وتركز على تقنيات التعلم الالي والذكاء الاصطناعي وتطبيقاتها والاتجاهات المستقبلية في الكيمياء العضوية.

Abbreviations:

ML: Machine Learning.

AI: Artificial Intelligence.

DL: Deep Learning.

FNN: Feed Forward Neural Networks.

MLP: Multi-Layer Perceptron's.

CNN: Convolutional Neural Networks.

RNN: Recurrent Neural Networks.

LSTM: Long Short-Term Memory.

GAN: Generative Adversarial Networks.

RL: Reinforcement Learning.

DP: Dynamic Programming.

SARSA: State-Action-Reward-State-Action.

TD: Temporal Difference.

DQN: Deep Q-Learning.

ADMET: Absorption, Distribution, Metabolism, Excretion, Toxicity.

VS: Virtual Screening.

TCGA: The Cancer Genome Atlas.

Figure: **Page:**

Scheme: **Page:**

1.Introduction

1.1 Overview of ML and AI in Organic Chemistry

Artificial intelligence and machine learning have emerged as important topics in recent years, especially in the field of study and education, as these technologies enhance the teaching and learning process. In the field of chemistry education, there is increasing interest in integrating these technologies into the curriculum, including laboratory experiments, simulations, virtual laboratories, data analysis, and assessment.[1] However, the available literature on the status of artificial intelligence and machine learning in chemistry education is still limited.[2] The text discusses the integration of artificial intelligence and machine learning into organic chemistry education, which includes the use of intelligent teaching systems, virtual laboratories, and personalized learning platforms.[3] These technologies rely on machine learning algorithms to adapt education to the needs of all scientists and student, providing personalized instruction and feedback.[4] Compared to traditional teaching, these methods provide interactive learning experiences that help students build their knowledge more effectively.[5] They also allow us to access a personalized learning environment that adapts to our learning style and provides immediate feedback.[6] In addition, these technologies facilitate simulations and virtual laboratory experiments, allowing scientists to observe chemical reactions in ways that are not possible in traditional laboratories.[7] Machine learning techniques are increasingly being used in chemistry to gain deeper insights into chemical systems or processes, reducing the need for time, computational resources, and physical materials.[8] These techniques can help collect repetitive information faster, enabling the discovery of underlying patterns that provide far more chemical insights with less data and experiments than traditional studies, and may also contribute to the creation of new insights that were not previously possible.[9] The research explores the relationship between chemistry and artificial intelligence (AI), noting that chemistry is often understood through the interactions of new chemicals and drugs.[10] However, AI has become an integral part of modern chemistry, being used in the design of synthesis systems, the interpretation of spectral data, and the modeling of molecules.[11] In this review, we will focus mainly on the Techniques and applications of AI in organic chemical synthesis, illustrating the promising future directions for AI in chemistry.

2. Key Techniques in ML and AI for Organic Chemistry

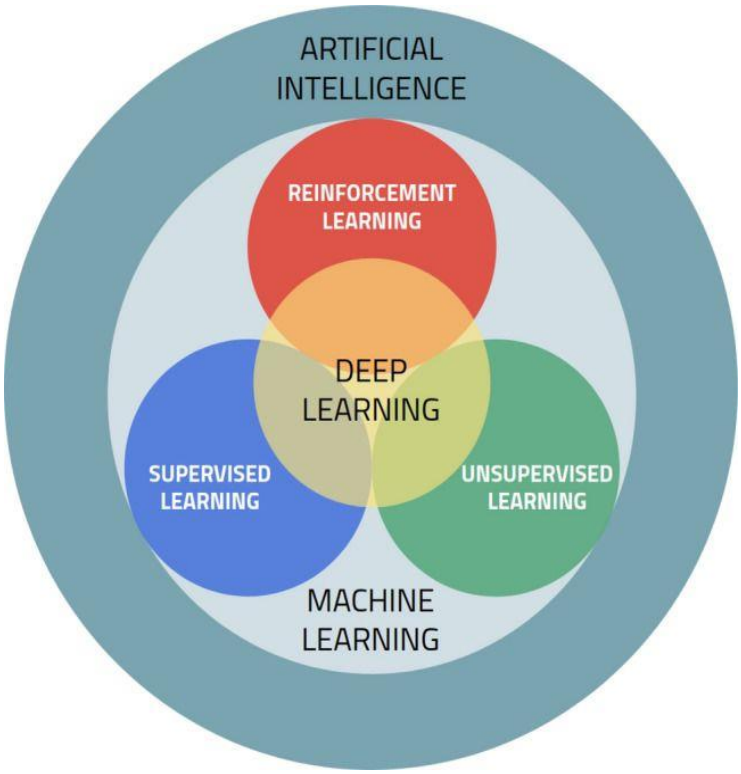


Figure 1: Key Techniques in ML and AI for Organic Chemistry.

2.1. Supervised Learning

Supervised learning is a method in machine learning where an algorithm is trained using data that includes both the inputs and their corresponding correct outputs. This process enables the algorithm to understand the relationship between the inputs and the desired outputs, allowing it to make accurate predictions on new, unseen data.[12] This technique is commonly applied in tasks such as classification, where the goal is to categorize data into different groups, and regression, where the aim is to predict continuous values, prediction of toxicity, solubility and pKa value[13].

2.2. Unsupervised Learning

Unsupervised learning involves the use of machine learning algorithms to understand patterns and structures in unlabeled data. Unlike supervised learning, there are no predefined outputs provided to guide the process. The algorithm independently identifies relationships and structures within the data. This technique is often employed in tasks like clustering, which involves grouping similar data points, and dimensionality reduction, which simplifies the data by reducing the number of variables or features. [14]

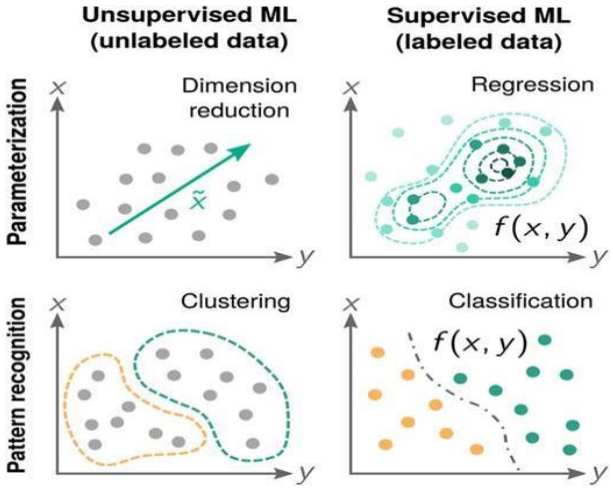


Figure 2: The difference between supervised ML and unsupervised ML[15]

The **Figure2** shows the different ML model categories. Unsupervised learning technique is often used for dimensionality reduction or clustering which obtained by unlabeled data, whereas supervised learning technique is used for regression or classification which obtained by labeled data.

2.3. Deep Learning (DL)

Deep Learning is a rapidly evolving branch of machine learning, focusing on advanced learning and prediction tasks, supported by a wide range of chemical literature, open-source software, and datasets. The ability of deep learning to identify relevant phenomena, accelerate chemical reactions, predict important properties, determine

optimal synthesis routes, resolve critical analytical uncertainties, and reduce costs and resources offers significant benefits in the field of chemistry.[16] Its success in modeling compound properties and reactions depends on access to comprehensive repositories of published chemical data.[17] However, there are barriers to overcome, such as data cleaning, generating accurate and meaningful chemical information, lack of standardization of chemical data, and lack of expertise and knowledge of machine and deep learning in the chemistry sectors. [18]

2.3.1. Feed forward neural networks (FNNs)

also known as multi-layer perceptron's (MLPs), represent the most basic and simple form of deep learning.[19] These networks operate by flowing information in one direction from input layers to output layers, without any feedback loops.[20] FNNs are widely used in tasks such as image classification, text analysis, and regression problems.[21]

2.3.2. Convolutional Neural Networks (CNNs)

very popular because of their exceptional performance in image and video tasks.[22] These networks have specialized convolutional layers that enable them to discover patterns and features within images, making them effective in recognizing objects, shapes, and textures.[23]

2.3.3. Recurrent Neural Networks (RNNs)

Handle sequential data such as time series and linguistic information, and feature loops that help retain information over time, making them effective for tasks such as speech recognition, language modeling, and translation.[24]

2.3.4. Long Short-Term Memory (LSTM)

Networks are an advanced type of RNN, specifically designed to address the gradient leakage problem, giving them a better ability to capture long-term dependencies in sequential data, providing valuable solutions in multiple applications.[25]

2.3.5. Generative Adversarial Networks (GANs)

Are an exciting pair of neural networks, consisting of a generative network and a discriminative network, trained together in a competitive process. [26] GANs excel at generating realistic synthetic data, such as images, audio, and text, making them a powerful tool in creative fields.[27]

2.4. Reinforcement Learning (RL)

RL includes an agent learning excellent strategies through test and error by combining with an ecosystem to increase additive rewards. A fundamental challenge is the analyze-exploit problem, where the agent must balance exploiting actions with known rewards and exploring new, untested actions to determine potentially better results. Machine learning combines various algorithms to enable systems to learn and improve through data analysis. Hybrid approaches like semi-supervised learning and transfer learning blend paradigms, using labeled and unlabeled data or transferring knowledge from one task to another for enhanced performance. [28]

2.4.1. Q-Learning

This is model-free reinforcement learning. It can also be viewed as a method of asynchronous dynamic programming (DP). The value-based reinforcement learning algorithm is used to find the optimal action-selection policy in each environment. Explores the best chemical pathways by learning from trial and error.[29]

2.4.2. State-Action-Reward-State-Action (SARSA)

Is an on-procedure reinforcement learning algorithm used to learn the prime procedure for a given environment. It is a kind of temporal difference (TD) learning technique. The fundamental idea of SARSA is that the agent learns from the developments of its actions by interacting with the environment and updating its policy based on the experiences it congregates. [30]

2.4.3. Deep Q-Learning (DQN)

Deep Q-learning is an extension of Q-learning, a popular model-free reinforcement learning algorithm, that uses deep neural networks to approximate the Q-value function. It is particularly useful for solving problems with large or continuous state spaces where traditional tabular Q-learning becomes infeasible such as video games.[31]

3. Applications of ML and AI in Organic Chemistry:

3.1 Molecular Property Prediction

Refers to the process of predicting the physical, chemical, or biological properties of molecules based on their structure, using computational models, machine learning algorithms, or quantum mechanical calculations. These properties can include a wide range of characteristics, such as solubility, boiling point, toxicity, molecular weight, polarity, electronic structure, and bioactivity, which are crucial in areas like drug discovery, materials science, and chemical engineering. [32] Drug Activity Prediction, Toxicity, and ADMET (Absorption, Distribution, Metabolism, Excretion, Toxicity) In drug discovery, predicting the molecular properties of compounds is crucial for identifying lead candidates that are likely to have favorable pharmacokinetics and minimal toxicity.[33] By using molecular property prediction, researchers can assess how

well a drug candidate will interact with its target, its potential side effects, and its ability to be absorbed, distributed, metabolized, and excreted by the body.[34] Predicting the toxicity of novel compounds is critical to avoid adverse effects in clinical trials. Machine learning models have been developed to predict the toxicity of drug candidates before they are synthesized and tested in the lab. [35] Predicting Environmental Impact, Biodegradability, and Toxicity In environmental science, molecular property prediction is used to assess the potential environmental impact of chemical pollutants. It helps in evaluating the biodegradability, persistence, and toxicity of chemicals, thus guiding the development of safer chemicals and remediation strategies. [36] Large databases of known compounds and their properties can now be explored using machine learning techniques, which then use the data to suggest new possibilities, as shown in **Figure 3**

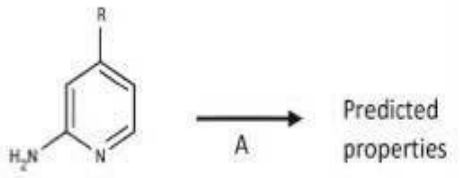


Figure 3: AI application predicting chemical properties.

3.1.1 Docking:

ML enhances computational docking by improving scoring functions that predict ligand-protein binding stability. This accelerates drug candidate identification and testing, revolutionizing computer-aided drug design.[37] Molecular docking predicts the optimal orientation and conformation of molecules in protein active sites. Scoring functions evaluate binding affinity and select the best complex, making docking a valuable tool in drug design. [38] Docking is essential in drug design but faces challenges like incorrect binding sites, unsuitable databases, and unreliable scores the need for caution to ensure accurate results. [39] Molecular docking predicts ligand binding modes to proteins, aiding drug design and lead optimization. It involves efficient searches, accurate scoring, and careful input setup.[40] Glide, a popular docking program, uses hierarchical filters to explore ligand poses and orientations in receptor binding sites, incorporating ligand flexibility through exhaustive torsion angle searches. [41]

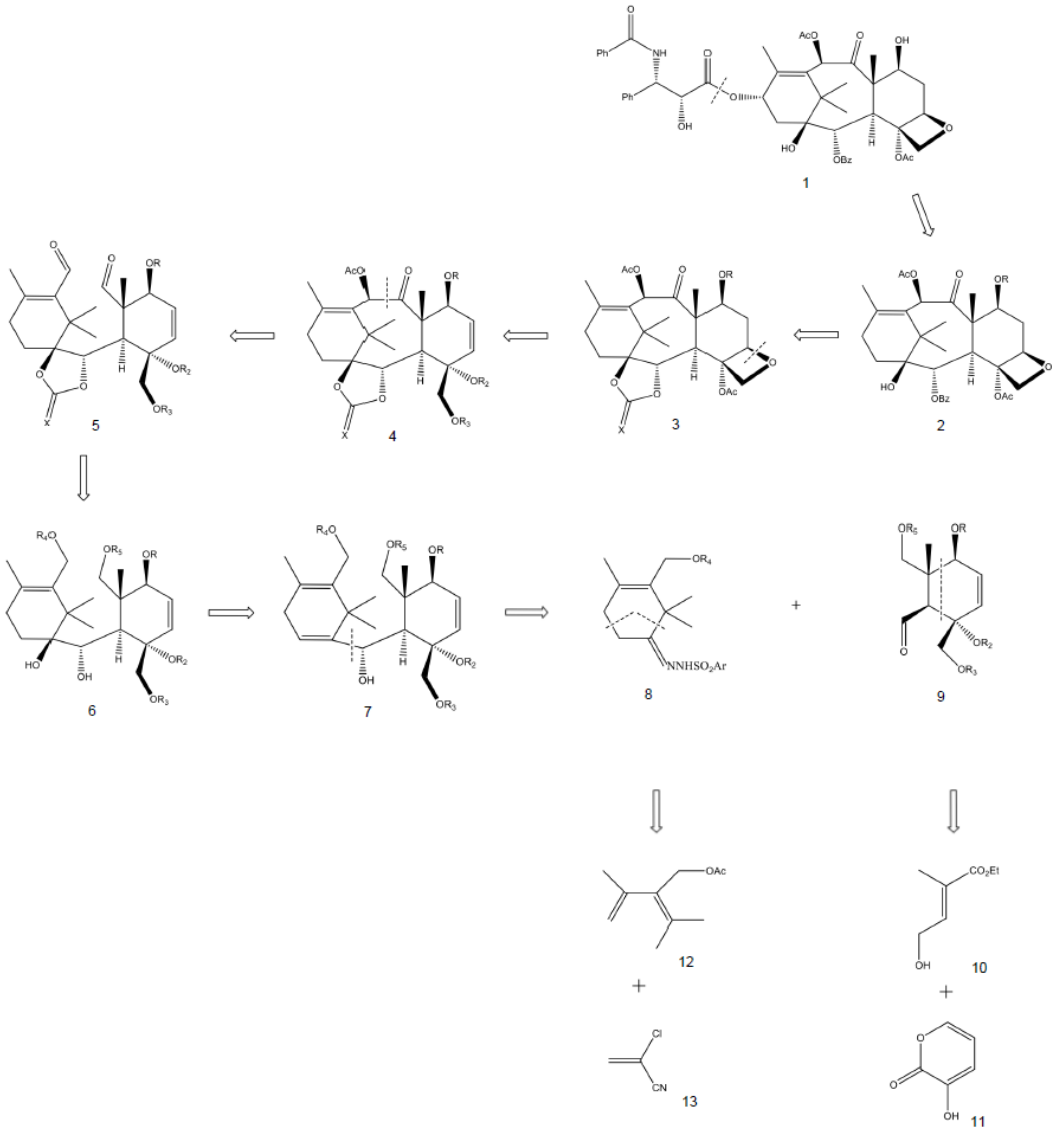
3.2 Chemical Reaction Prediction

Is the process of forecasting the products and reaction pathways of a chemical reaction based on a given set of reactants, reaction conditions (such as temperature, pressure, and solvent), and sometimes catalysts.[42] It involves applying computational methods, including machine learning, quantum chemistry, and data-driven approaches, to predict how molecules will interact, break bonds, and form new ones.[43] The goal is to predict the outcome of a reaction accurately without physically conducting the experiment, enabling more efficient and cost-effective chemical synthesis.[44] Reaction Outcome Prediction, Catalyst Efficiency In chemistry, particularly in catalysis, molecular property prediction helps in predicting the outcome of chemical reactions, including the selectivity of a reaction and the identification of potential catalysts. [45] Predicting how a molecule will behave in a reaction allows chemists to design more efficient catalysts, reducing the need for trial-and-error experimentation.[46] In the pharmaceutical industry, predicting chemical reactions is essential for the development of new drugs and materials. [47] By predicting reaction pathways and intermediates, researchers can efficiently design and synthesize complex molecules with desired pharmacological properties.[48] Chemical reaction prediction plays a vital role in the discovery of novel materials, such as polymers, catalysts, and superconductors.[49] Predicting reactions allows researchers to identify new material compositions and optimize existing ones.[50] In environmental chemistry, predicting chemical reactions can help design processes for the treatment of pollutants or the recycling of waste products.[51] This is essential in reducing the environmental footprint of industrial processes and managing hazardous chemicals.[52] n biochemistry, chemical reaction prediction plays a key role in understanding and designing metabolic pathways, enzyme mechanisms, and protein-ligand interactions.[53] This is important for the development of biocatalysts, biosensors, and synthetic biology applications.[54] Example: Automated Reaction Prediction in Drug Discovery Software tools like **ROSIE** and Chematica are used in pharmaceutical research to predict reaction pathways and suggest the best synthetic routes for drug candidates. For instance, when developing a new antibiotic or anticancer drug, these systems can propose the most efficient and cost- effective methods to synthesize the compound.[55]

3.3 Retrosynthesis and Synthesis Planning

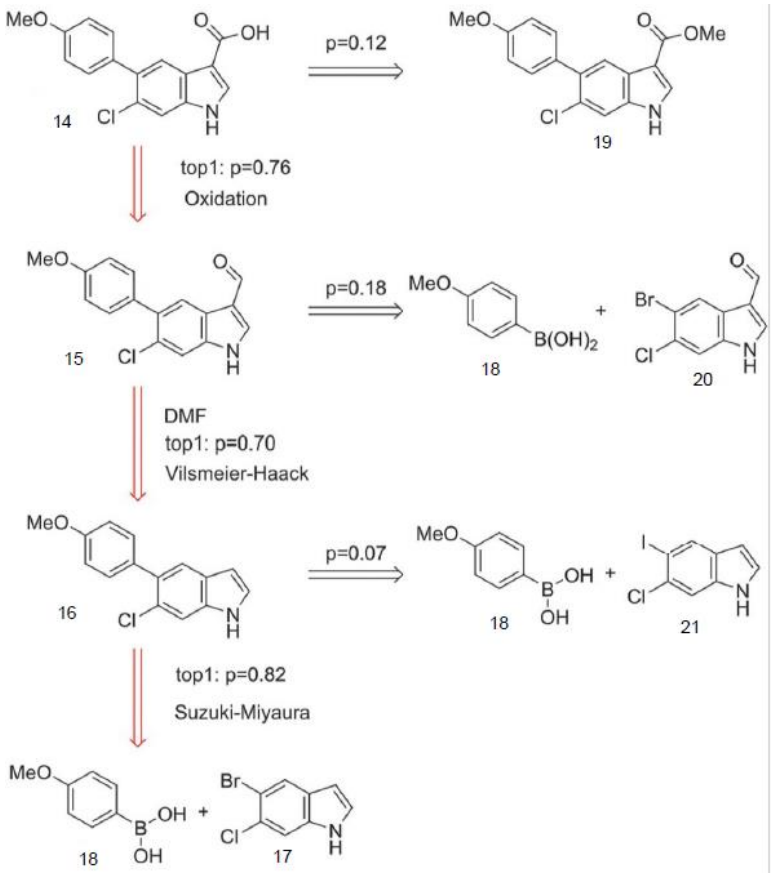
Retrosynthesis, a common technique in organic chemistry synthesis planning, involves breaking down a molecule step-by-step into smaller subunits until compounds that are easy to synthesize, or purchase are identified.[56] The ultimate objective is to create a synthesis roadmap for the target compound. With the aid of computer-assisted retrosynthesis, both chemists and machine learning researchers aim to speed up chemical synthesis, conserving time and resources, tackling more complex molecules, and generating more efficient and safer routes. These synthesized pathways can be invaluable for medical chemists in creating molecules of interest,[57] forming the foundation for autonomous chemistry,[58] or integrating into de novo drug design to evaluate synthesizability.[59] In recent years, retrosynthesis prediction has gained popularity due to its alignment with modern machine learning techniques, enabling consideration of a broader range of synthesis routes. This field is divided into two areas: single-step retrosynthesis prediction, which breaks down a product into a single set of reactants, and multi-step synthesis planning, which uses search algorithms to find synthesis routes to commercially available compounds.[60]

Retrosynthesis is used in drug discovery to identify efficient synthetic routes for novel antibiotics, also retrosynthesis plays a critical role in designing drugs that are capable of target cancerous cells. The anticancer drug Taxol **(Scheme 1)** was synthesized using retrosynthetic methods to create an efficient process for large-scale production. The process involves complex organic reactions to assemble the core structure, starting from simple molecules.[61]



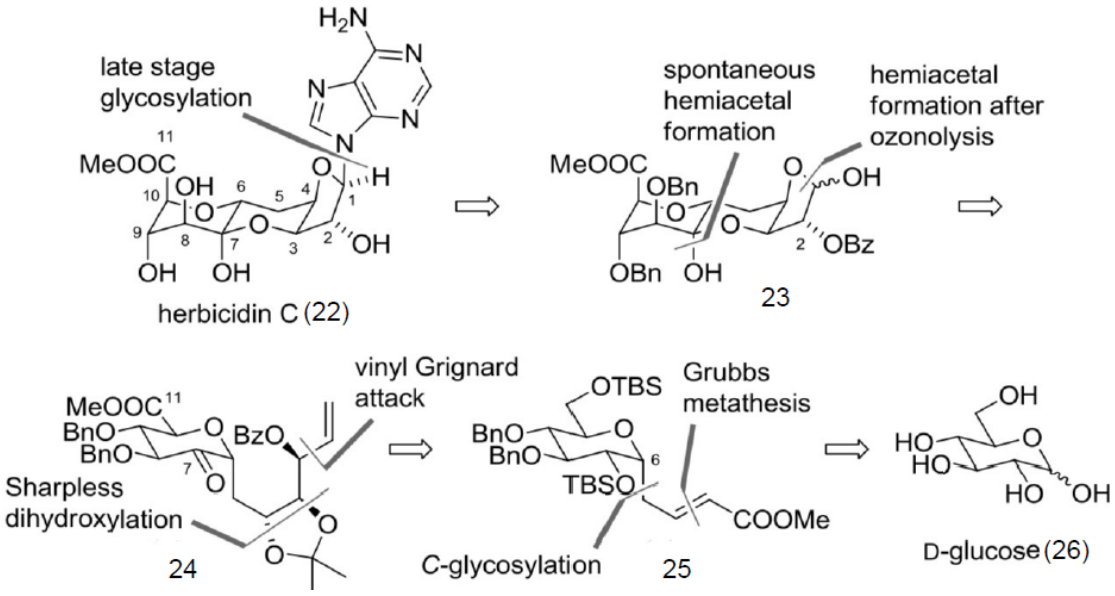
Scheme 1: Retrosynthetic Analysis of Taxol.

Also, Retrosynthetic used for analysis of 6-Chloro-5-(4-methoxyphenyl)-1H-indole-3-carboxylic acid a drug molecule **(Scheme 2)**.[62]



Scheme 2: Retrosynthetic analysis of 6-Chloro-5-(4-methoxyphenyl)-1H-indole-3-carboxylic acid a drug molecule.

Retrosynthesis is used as well to design effective herbicides with novel mechanisms of action and improved environmental profiles. Such as herbicides for use in agriculture. Due to their diverse biological activity **(Scheme 3)**.[63] Retrosynthesis provided many examples and guidelines that helped in organic synthesis protocols using biocatalysts.[64] Computer-aided retrosynthesis has already shown it can create practical routes for making moderately complex targets. Which is possible to find effective, eco-friendly, and experimentally feasible total synthetic pathways for even more complex synthetic routes often found in active pharmaceutical ingredients.[65]

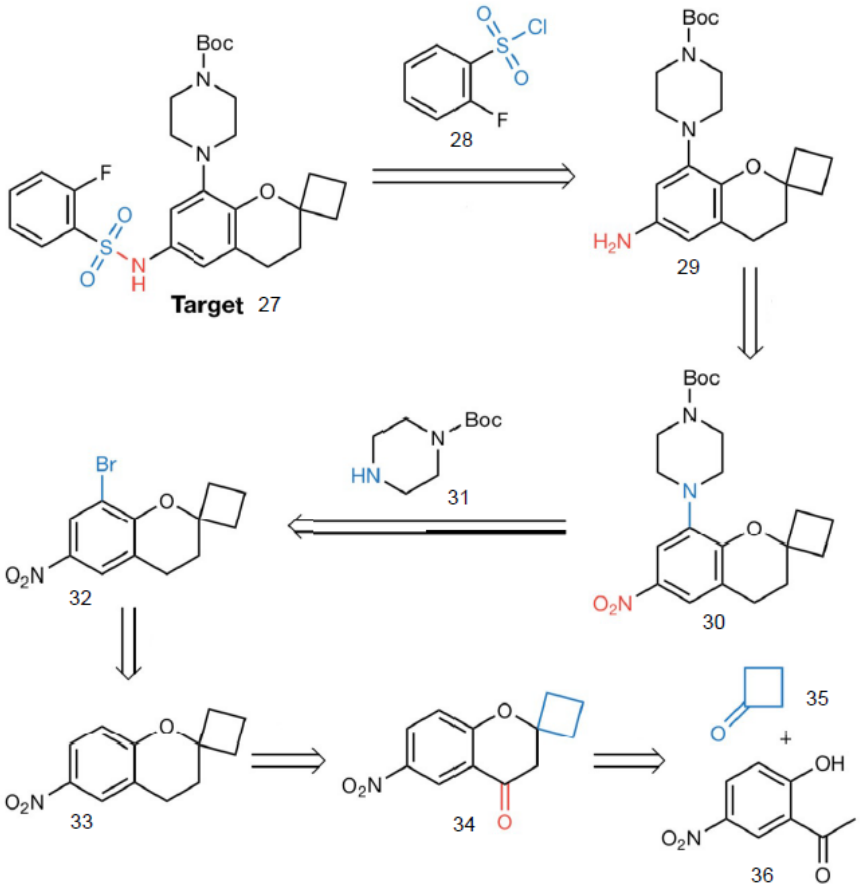


Scheme 3: An example of using Retrosynthetic analysis for the left side approach of herbicidin C.

3.4. Drug Discovery

Over the past two decades, drug discovery has faced challenges like inefficiency, high costs, and time-intensive processes. [66] Traditional computational methods, such as virtual screening (VS) and molecular docking, have helped but remain limited by inaccuracy and inefficiency.[67] Artificial intelligence (AI), incorporating machine learning (ML) and deep learning (DL), has emerged as a transformative solution.[68] AI enhances drug discovery by streamlining complex steps, including target identification, compound screening, lead optimization, and evaluating drug toxicity and efficacy. Additionally, AI reduces costs and accelerates timelines by handling massive chemical libraries and minimizing off-target toxicity.[69] The digitization of pharmaceutical and healthcare data has further motivated AI integration, enabling efficient analysis of complex datasets.[70] Applications of AI in drug discovery include primary and secondary screening, drug monitoring, dosage optimization, drug repositioning, and

poly pharmacology. This highlights how AI, alongside conventional chemistry, revolutionizes drug development by improving efficiency, reducing costs, and overcoming traditional limitations. [71] Machine learning has greatly enhanced virtual screening, where AI algorithms sift through large compound databases to identify promising drug candidates by predicting their biological activity, interaction with targets, and ADMET (Absorption, Distribution, Metabolism, Excretion, and Toxicity) properties. [72] AI models are also increasingly used in designing peptide-based drugs and protein therapeutics, where the challenge is to predict the structure and function of peptides and proteins that can act as effective therapeutics.[73] researchers used AI to design novel peptides that could bind to specific receptors in the body, offering new opportunities for peptide-based therapies. [74] AI and ML have been applied to genomic and transcriptomic data to identify new drug targets, particularly in cancer and rare diseases. [75] Machine learning models analyze data from sources like The Cancer Genome Atlas (TCGA) or gene expression databases to identify genetic mutations that could serve as therapeutic targets. [76] Driver ML is a machine learning tool developed for identifying driver genes in cancer, which is a crucial step in cancer drug discovery. It analyzes genomic data to find genes that are critical for tumor development. [77] In 2020, AI models were used to identify existing drugs that could be repurposed for the treatment of COVID-19. The AI platform screened millions of compounds and suggested drugs that could target the virus. [78]



Scheme 4: drug candidate synthesis.

An example tert-butyl 4-[6-(2-fluorobenzenesulfonamido)-3,4-dihydrospiro[1-benzopyran-2,1'-cyclobutane]-8-yl]cyclohexane-1-carboxylate is the synthesis of an exemplary six-step synthesis route for an intermediate in a drug candidate synthesis. The affected functional groups in each step are marked blue or red. shows an exemplary six-step route for an intermediate of a drug candidate synthesis reported in 2015, which was found by our algorithm in 5.4 s. Several hundred additional exemplary retrosynthetic routes found by the MCTS algorithm for molecules first synthesized in or after 2015 are deposited in Supplementary Information.[79]

4. Experimental:

Although this is a literature review, we managed to conduct a small test regarding retrosynthesis. We set the compound **37 (Figure 4)** as a target and then we started using the available sources to perform the retrosynthesis and develop the synthetic plan.

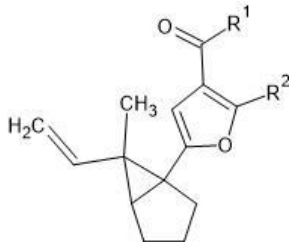


Figure 4 : target compond 37

4.1 Retrosynthesis:

The retrosynthesis starts by converting diene **38** into alkene **37** through a cyclization reaction. The diene **38** can be obtained from alkyne **39** by formalization, followed by a Knoevenagel condensation reaction to generate diene **38**. Alkyne **39** can be synthesized from ester **40** by the reduction of the ester, followed by protection of the alcohol. Finally, ester **40** can be produced from alcohol **41** by oxidation, followed by a Wittig reaction to yield ester **40**.

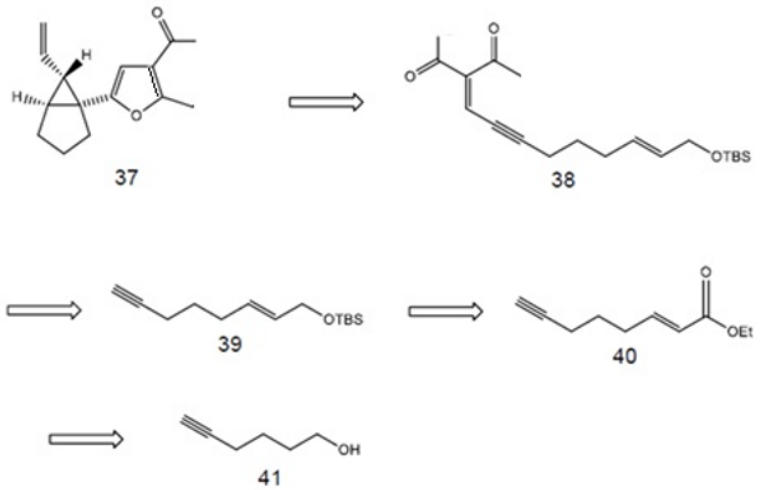


Figure 5: The retrosynthesis of alkene 37

This retrosynthesis was taken from a reference [86] through sic-finder which rely on real data and provide us with the reactions condition and steps. We can customize the starting materials to our needs by adding a methyl group to carbon 2 and changing the diketone in the Knoevenagel condensation step.

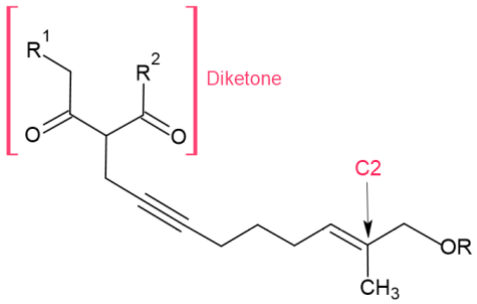


Figure 6: The position of the desire target

Another way was used is the RetroScheme software which give us idea of the possible reaction to reach the desired target from different aspect of sicfinder, it does not provide reference or condition it's just give an idea of possible conversion without more details. (Figure 5 and 6)

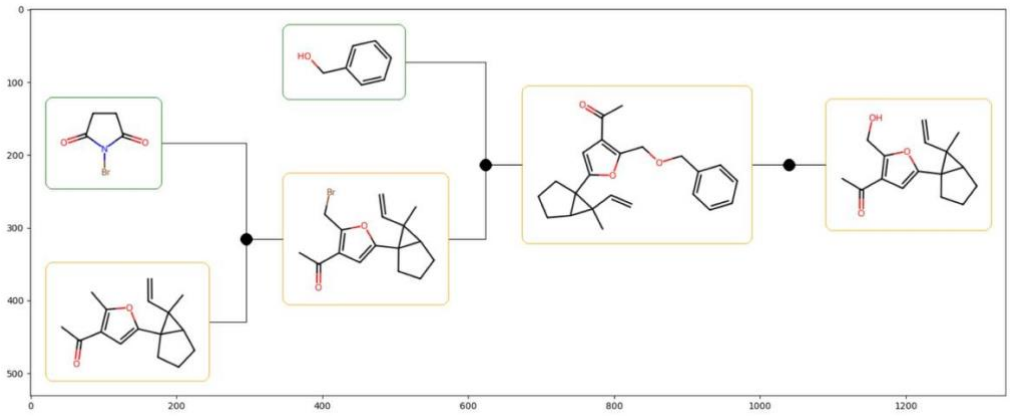


Figure 7: the first possible retrosynthesis using RetroShceme.

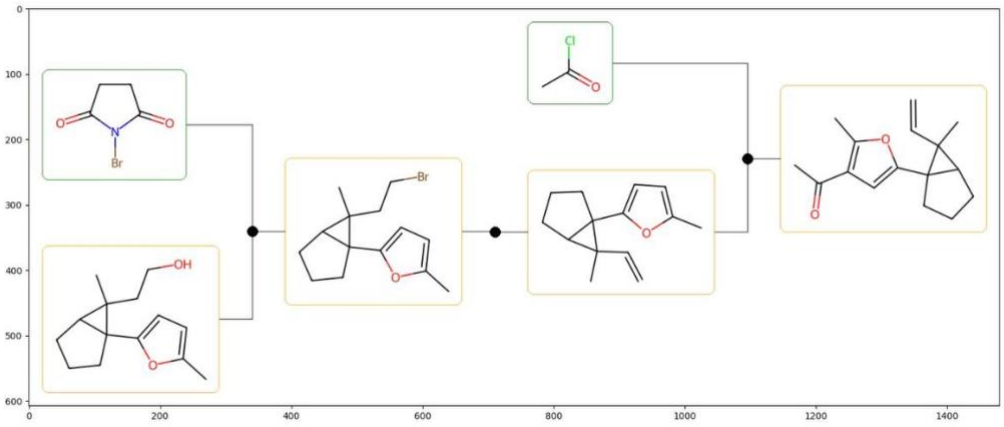


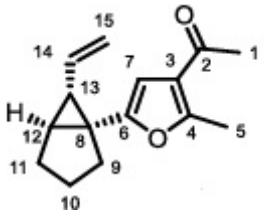
Figure 8: the first possible retrosynthesis using RetroShceme.

So the Al help us to find the route to synthesise alkene **37** regarding to real data using Sicfinder. There are many good paid sources can be used like Reaxys and many other software which proved a very good retrosynthesis plan.

4.2 NMR

The Al an ML can also help us to predict the NMR for our product before we synthesis them. Which give us an idea of how our products should look like. It not necessary to be the exact NMR but at least near to it, and then we could integrate our NMR and camper it with prediction of the software was used to predict.

Here we used the MasterNova to predict the NMR for compound **37** which was already synthesised [86]. Then we will compare our prediction to the real data. As shown in (Figure 9) and (Figure 10) the alkene **37** was paste in MasterNova and the NMR was predicted for H^1 and C^{13} NMR. The data we collect bretty match the data was provide by Varena.[86]



Varena group data ((1H NMR (400 MHz, CDCl3) δ 6.25 (1H, s, CH C7), 5.39 (1H, ddd, J = 17.1, 10.2, 9.3 Hz, CH-C14), 5.08 (1H, dd, J = 17.1, 1.8 Hz, CHH-C15trans), 4.88 (1H, dd, J = 10.2, 1.8 Hz, CHH-C15cis), 2.54 (3H, s, CH3-C5), 2.37 (3H, s, CH3-C1), 2.16 (1H, dd, J = 12.7, 8.4 Hz, CHH-C9), 1.94 (1H, dd, J = 12.7, 8.4 Hz, CHH-C9), 1.93–1.88 (2H, m, CH2-C11), 1.80–1.78 (1H, m, CH-C13), 1.77–1.71 (1H, m, CHH-C10), 1.75 (1H, dd, J = 4.1, 9.1 Hz, CH-C12), 1.43–1.32 (1H, m, CHH C10); 13C NMR (126 MHz, CDCl3) δ 194.5 (C-C2), 157.1 (C-C4), 154.4 (C-C6), 137.1 (CH-C14), 122.1 (C-C3), 113.9 (CH2-C15), 106.5 (CH-C7), 33.4 (C-C8), 32.7 (CH2-C9), 32.4 (CH-C12), 30.2 (CH-C13), 29.3 (CH3-C1), 27.5 (CH2-C11), 21.7 (CH2-C10), 14.6 (CH3-C5).

MasterNova prediction data 1H NMR (500 MHz, Chloroform) δ 6.02 (s, 1H, CH C7), 5.79 (ddd, *J* = 16.6, 10.0, 6.2 Hz, 1H, CH-C14), 4.97 (dd, *J* = 17.4, 1.4 Hz, 1H, CHH-C15 trans), 4.92 (dd, *J* = 10.4, 1.4 Hz, 1H, CHH-C15 cis), 3.65 (t, *J* = 6.5 Hz, 1H, CH-C13), 2.55 (s, 3H, CH3-C1), 2.20 (s, 3H, CH3-C5), 2.14 (dt, *J* = 6.8, 3.3 Hz, 1H, CH-C12), 1.82 – 1.75 (m, 1H, CHH-C9), 1.71 – 1.64 (m, 1H, CHH-C9), 1.53 – 1.36 (m, 4H, CH2-C10, CH2-C11). 13C NMR (125 MHz, Common NMR Solvents) δ 197.44 (C-C2), 168.68 (C-C4), 161.56 (C-C6), 134.91 (CH-C14), 123.69 (C-C3), 116.99 (CH2-C15), 115.68 (CH-C7), 39.52 (C-C8), 35.53 (CH2-C9), 28.64 (CH-C12), 28.16 (CH-C13), 24.26 (CH£-C1), 23.11 (CH2-C11), 18.96 (CH2-C10), 14.73 (CH3-C5).

The data match each other with slightly different in NMR shift. However the hydrogen on C13 did show different it shielded in the prediction data and shows at 3.65 shift with triplet where as in real data it shows at 1.80 shift with multiple. For this reason we understood that, the AI did not give a very accurate result but it help us to estimate the data and adjust it to our need.

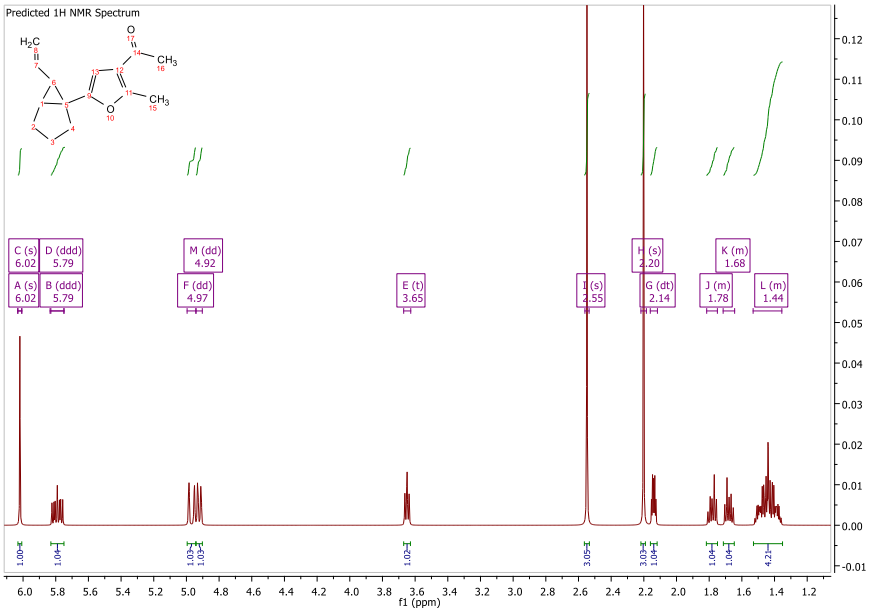


Figure 9: The H^1 NMR prediction for compound 37

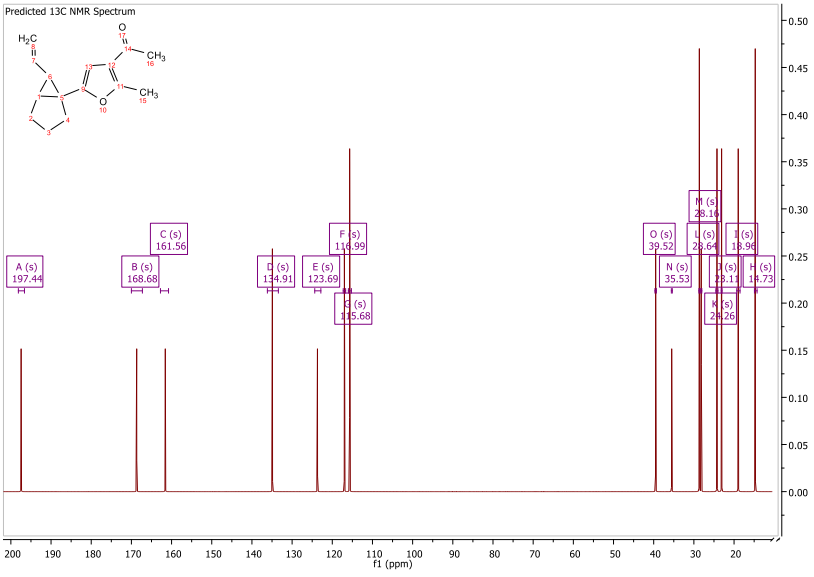


Figure 10: The C^{13} NMR prediction for compound 37

5. Challenges and Limitations

Although there are many exciting possibilities, there are still some technical challenges and limitations in using AI and ML in chemistry education. These challenges include the need for high-quality training data, powerful computers, and expertise in developing and using algorithms.[80] Additionally, AI models might have difficulty accurately interpreting complex chemical structures and reactions, requiring continuous improvement and validation by experts in the field.[81] Another challenge is the lack of regulations for using AI in education, which raises concerns among stakeholders about data protection and privacy.[82]

6. Future Directions

Synthetic organic chemistry is crucial for drug discovery, materials science, and chemical biology. However, it requires a lot of expertise and practice to carry out complex syntheses. Although attempts to automate chemical synthesis have been ongoing for decades with limited success, advancements in AI—powered by better computing, data, and algorithms—are changing the field. This review highlights recent AI-driven progress in synthetic chemistry.[83] This integration could lead to the development of automated laboratories.[84] The rapidly growing field of computational science, driven by advanced computing, data sharing, and open-source tools, has the potential to overcome many challenges, leading to a chemical discovery revolution.[85]

7. Conclusion

Organic chemistry is moving from hard manual work to a time driven by automation and AI, thanks to technological progress and the need for efficiency and accuracy. AI models are transforming how we plan synthetic processes, while autonomous robotic systems speed up discoveries with precision. This work aims to explain the basic concepts of AI to help laboratory chemists see it as a useful tool instead of fearing it as a competitor, by identifying knowledge gaps for future research and showing how chemical AI will operate in the digital chemistry era.

References:

1- A.Lyamuremye, F.N. Niyonzima, J. Mukiza, I. Twagilimana, P. Nyirahabimana, T. Nsengimana, J.D. Habiyaremye, O. Habimana, E. Nsabayezu. *Discover Education.* **2024**,3,95.

2- L. He, L. Bai, D.D. Dionysiou, Z. Wei, R. Spinney, C. Chu, Z. Lin, R. Xiao. *Chemical Engineering Journal.* **2021**,426,131810.

3- W.K. Chiu, *Education Sciences.* **2021**,11,709.

4- R. Lamb, K. Neumann, K. A. Linder. *Computers and Education Artificial Intelligence.* **2022**,3.100078.

5- H. K. Wu, P. Shah. *Science Education*. **2004**,88,465-492.

6- M. Cevikbas, G. *Kaiser. Sustainability.* **2022**,14,11393.

7- X. Jiang, S. Luo, K. Liao, J. Ma, J. Jiang, Z. Shuai. *Cell Reports Physical Science.* **2024**,5.102049.

8- M. Meuwly. *Chemical Reviews.* **2021**,121,10218-10239.

9- R. Houhou, T. Bocklitz. *Analytical Science Advances.* **2021**,2,128-141.

10- N. Brown, P. Ertl, R. Lewis, T. Luksch, D. Reker, N. *Schneider. Journal of Computer Aided Molecular Design.* **2020**,34,709-7015.

11- A.Y. Sharma, R. Sahni, S. Herald, R. Genius. *International Journal for Research in Applied Science & Engineering Technology*. **2019**,7,54-60.

12- R. S. A. E. Ali, J. Meng, M. E. I. Khan, X. Jiang. *Artificial Intelligence Chemistry*. **2024**, 2, 100049.

13- K. Hansen, G. Montavon, F. Biegler, S. Fazli, M. Rupp, M. Scheffler, O. A. V. Lilienfeld, A. Tkatchenko, K. R. Muller. *Journal of chemical theory and computation*. **2013**, 9, 3404-3419.

14- A. Karthikeyan, U. D. Priyakumar. *Journal of Chemical Sciences*. **2022**, 134, 1-20.

15- J. Westermayr, M. Gastegger, K. T. Schutt, R. J. Maurer. The *Journal of Chemical Physics*. **2012**, 154.

16- F.G.G. Tania, A.C.C. Alberto. *Frontiers in chemistry*. **2019**,7,809.

17- G.B. Goh, N.O. Hodas, A. Vishnu. *Journal of computational chemistry*. **2017**,38,1291-1307.

18- C. Trinh, D. Meimaroglou, S. Hoppe. *Processes*. **2021**,9,1456.

19- Z. Zhang, F. Feng, T. Huang. *Applied Sciences*. **2022**,12,12478.

20- M.T. Hagan, H.B. Demuth, O.D. Jesus. *International Journal of Robust and Nonlinear Control.* **2002**,12,959-985.

21- A.H. Blanco, B.H. Flores, D. Tomas, B.N. Colorado. *Complexity*. **2019**,1306039.

22- J.Gu, Z.Wang, J.Kuen, L.Ma, A.Shahroudy, B.Shuai, T.Liu, X.Wang, G.Wang, J.Cai, T.Chen. *Pattern recognition*. **2018**,77,354,377.

23- I.Namatevs. *Information Technology and Management Science*. **2017**,20,40-47.

24- S. Grossberg. *Scholarpedia*. **2013**,8,1888.

25- F. Kratzert, D. Klotz, C. Brenner, K. Schulz, M. Herrnegger. *Hydrology and Earth System Sciences*. **2018**,22,6005-6022.

26- H.Alqahtani, M.K.Thorne, G.Kumar. *Archives of Computational Methods in Engineering*. **2019**,28,525,552.

27- T. Chakraborty, U. Reddy K S, S.M. Naik, M. Panja, B. Manvitha. *Machine Learning: Science and Technology*. **2024,**5,011001.

28- R. S. A. E. Ali, J. Meng, M. E. I. Khan, X. Jiang, *Artificial Intelligence Chemistry,***2024**,2,100049.

29- C. J. C. H. Watkins, P. Dayan, *Machine Learning,* **1992**, 8, 279-292

30- D. Zhao, H. Wang, K. Shao, Y. Zhu, *IEEE symposium series on computational intelligence (SSCI),* **2016**,1-6.

31- F. Tan, P. Yan, X. Guan, *Neural Information Processing,* **2017**, 475-483.

32- J.Shen, C.A. Nicolaou. *Drug Discovery Today*: *Technologies*. **2019**,32,33.

33- J. Lin, D.C. Sahakian, S.M. De Morais, J.J. Xu, R.J. Polzer, S.M. Winter. *Current topics in medicinal chemistry*. **2003**,3,1125-1154.

34- V.G. Maltarollo, J.C. Gertrudes, P.R. Oliveira, K,M, Honorio. *Expert opinion on drug metabolism & toxicology*. **2015**,11,259-271.

35- L. John, H.J. Mahanta, Y. Soujanya, G.N. Sastry. *Computers in biology and medicine*. **2023**,153,106494.

36- H. Yin, C. Lin, Y. Tian, A, Yan. *Chemical Research in Toxicology*. **2023**,36,617-629.

37- M. A. Khamis, W. Gomaa, W. F. Ahmed, *Artificial intelligence in medicine,* **2015,** 63,135-152.

38- Z. Liu, Y. Liu, G. Zeng, B. Shao, M. Chen, Z. Li, Y. Liu, Y. Zhang, H. Zhong, *Chemosphere,* **2018,** 203,139-150.

39- Y. C. Chen, *Trends in pharmacological sciences,* **2015,** 36,78-95.

40- G. M. Morris, M. L. Wilby, *Molecular modeling of proteins,* **2008,** 365-382.

41- X. Y. Meng, H. X. Zhang, M. Mezei, M. Cui, *Current computer-aided drug design,* **2011,** 7,146-157.

42- W.A. Warr. *Molecular Informatics*. **2014**,33,469-476.

43- M.A. Kayala, P. Baldi. *Journal of chemical information and modeling*. **2012**,52,25262540.

44- V. Venkatasubramanian, V. Mann. *Current Opinion in Chemical Engineering*. **2202**,36.100749.

45- C.W. Coley, R. Barzilay, T.S. Jaakkola, W.H. Green, K.F. Jensen. *ACS central science*. **2017**,3,434-443.

46- L.G. Otten, F. Hollmann, I.W.C.E. Arends. *Trends in biotechnology*. **2010**,28,46-54.

47- K.R. Campos, P.J. Coleman, J.C. Alvarez, S.D. Dreher, R.M. Garbaccio, N.K. Terrett, R.D. Tillyer, M.D. Truppo, E.R. Parmee. *Science*. **2019**,363,0805.

48- N.A. Frolov, A.N. Vereshchagin. ." *International Journal of Molecular Sciences*. **2023**,24,2937

49- Akinpelu, M. Bhullar, Y. Yao. *Journal of Physics*: *Condensed Matter*. **2024**,36,453001.

50- T. Zhou, Z. Song, K. Sundmacher. *Engineering*. **2019**,5,1017-1026.

51- M.Behera, J.Nayak, S.Banerjee, S.Chakrabortty, S.K.Tripathy. *Journal of Environmental Chemical Engineering*. **2021**,9,4.

52- W.P.R.T. Perera, R.T. Hewage, H. Gunathilaka, J.A. Liyanage. *Integrated Science for Sustainable Development Goal 3*.**2024**,24, 201-243.

53- Dhakal, C. Mckay, J.J. Tanner, J. Cheng. *Briefings in Bioinformatics*. **2022**, 23, 476.

54- C.F. Moratti, C. Scott, N.V. Coleman. *Frontiers in Bioengineering and Biotechnology*. **2022**, 9, 804234.

55- S. Lyskov, F.C. Chou, S.O. Conchuir, B.S. Der, K. Drew, D. Kuroda, J. Xu, B.D. Weitzner, P.D. Renfrew, P. Sripakdeevong, B. Borgo, J.J. Havranek, B. Kuhlman, T. Kortemme, R. Bonneau, J.J. Gray, R. Das. *PloS one*. **2013**, 8. 63906.

56- M. H. S. Segler, M. Preuss, M. P. Waller. *Natuer*. **2018**, 555, 604-610.

57- C. W. Coley, W. H. Green, K. F. Jensen. *Accounts of chemical research*. **2018**, 51, 1281-1289.

58- C. W. Coley, D. A. Thmas III, J. A. M. Lummiss, J. N. Jaworski, C. P. Breen, V. Schultz, T. Hart, J. S. Fishman, L. rogers, H. gao, R. W. Hicklin, P. P. Plehiers, J. Byington, J. S. Piotti, W. H. Green, A. J. Hart, T. F. Jamison, K. F. Jensen. *Science.* **2019**, 356, 1566.

59- F. Miljkovic, R. Rodriguez-Perez, J. Bajorath. *ACS Omega*. **2021**, 6, 33293-33299.

60- P. Schwaller, A. C. Vaucher, R. Laplaza, C. Bunne, A. Krause, C. Corminboeuf, T. Laino. *Wiley Interdisciplinary Reviews: Computational Molecular Science*. **2022**, 16047.

61- K. C. Nicolaou, P. G. Nantermet, H. Ueno, R. K. Guy, E. A. Couladouros, E. J. Sorensen. *Journal of the American Chemical society*. **1995**, 117, 624-633.

62- M. H. S. Segler, M. P. Waller. *Chemistry--A European Journal*. **2017**, 23, 5966-5971.

63- D.hager. *Total Synthesis of Herbicidin C and Studies Toward Stephadiamine and Divergolides*. **2012**, 1-28.

64- R. O. M. A. De souza, L. S. M. Miranda, U. T. Bornscheuer. *Chemistry--A European Journal*. **2017**, 23, 12040-12063.

65- R. I. Teixeira, M. Andresini, R. Luisi, B. Benyahia. *JACS Au*. **2024**, 4, 4044-4543.

66- C. F. Lipinski, V. G. Maltarollo, P. R. Oliveira, A. B. F. Da silva, K. M. Honorio, *Frontiers in Robotics and AI*, **2019**,6,108.

67- P. Hassanzadeh, F. Atyabi , R. Dinarvand, *Advanced drug delivery reviews*, **2019**, 151,169-190.

68- W. Duch, K. Swaminathan, J. Meller, *Current pharmaceutical design*, **2007**,13,1497-1508.

69- L. Zhang, J. Tan, D. Han, H. Zhu, *Drug discovery today*, **2017**, 22,1680-1685.

70- A. M. Jordan, *ACS medicinal chemistry letters*, **2018**,9,1150-1152.

71- R. Gupta, D. Srivastava, M. Sahu, S. Tiwari, R. K. Ambasta, P. Kumar, *Molecular diversity*, **2021**, 25,1315-1360.

72- K. Swanson, P. Walther, J. Leitz, S. Mukherjee, J. C. Wu, R. V. Shivnaraine, J. Zou, *Bioinformatics*, **2024**, 40,416.

73- S. Hashemi, P. Vosough, S. Taghizadeh, A. Savardashtaki, *Heliyon*, **2024**, 10,40265.

74- M. Goles, A. Daza, G. C. Mora, L. S. Varon, J. S. Yanez, H. A. Kazemabad, M. D. Davari. R. U. Paredes, A. O. Nappa, M. A. Navarrete, D. M. Ortiz, *Briefings in Bioinformatics*, **2024**, 25,4.

75- S. Quazi, *Medical Oncology*, **2022**, 39,120.

76- M. Mohammed, H. Mwambi, I. B. Mboya, M. K. Elbashit, B. Omolo, *Scientific reports,* **2021**, 11,15626.

77- Y. Han, J. Yang, X. Qian, W. C. Cheng, S. H. Liu, X. Hua, L. Zhou, Y. Yang, Q. Wu, P. Liu, Y. Lu, *Nucleic acids research,* **2019**, 47, e45.

78- Y. Zhou, P. F. Wang, P. J. Tang, P. R. Nussinov, P. F. Cheng, *The Lancet Digital Health,* **2020**, 2, e667-e676.

79- M. H. S. Segler, M. Preuss, M. P. Waller, *Nature*, **2018**, 555, 604-610.

80- H. Abuhassna, F. Awae, M. A. B. M. Adnan, M. Daud, A. S. B. Almheiri, *International Journal of Information and Education Technology*, **2024**, 14,5.

81- S. Maestrales, X. Zhai, I. Touitou, Q. Baker, B. Schneider, J. Krajcik, *Journal of Science Education and Technology*, **2021**, 30,239-254.

82- I. Lee, B. Perret, *Proceedings of the AAAI conference on artificial intelligence*, **2022**, 36,12783-12791.

83- A. F. D. Almeida, R. Moreira, T. Rodrigues, *Nature Reviews Chemistry*,, **2019**, 3, 589-604.

84- A. Karthikeyan, U. D. Priyakumar, *Journal of Chemical Sciences*, **2022**, 134,1-20.

85- A. Tkatchenko, *Nature communications*, **2020**, 11,4125.

86- V. Klaus, S. Witmann, H. M. Senn, J. S. Clark. *Organic & Biomolecular Chemistry*. **2018**, 16, 3970-3982.