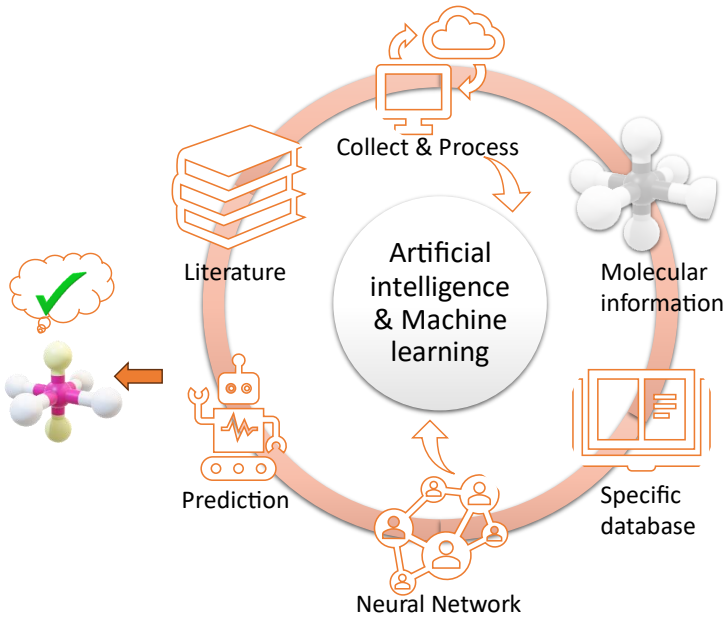
Exploring the Synergistic Potential of Artificial Intelligence and Machine Learning in Chemistry

Shyamal Mondal*

Department of Chemistry, New Alipore College, L-Block, New Alipore, Kolkata, West Bengal, India-700053

Date Received: 24-10-2024 Date Accepted: 25-12-2024



Abstract

Machine learning (ML) and artificial intelligence (AI) have become specialists in different areas of chemistry. These technologies help to change the standard approaches to data analysis and molecular design along with the properties forecast. This review describes the interesting applications and increasing potential of AI and ML, specifically in drug discovery, chemical synthesis, material science, and computational chemistry. Computationally, the focus was on the application of AI algorithms to quantum chemistry simulations to predict properties of elements within a molecule, or possible reactions of molecules at a rate that would not have been possible manually. Moreover, AI-driven robotic synthesis platforms and experimental techniques have become less labor-intensive. The methods used for the identification of new chemical structures have improved in terms of speed. The benefits and the limitations of integrating AI, as well as the opportunities, are discussed in detail. In this review, it is also reiterated that there are risks that come with the integration of ML in chemistry and how interdisciplinary collaboration and data sharing are crucial to advancing in this field. In a single summary, this review demonstrates how the use of AI and ML can and will expand the horizons of chemical science and discovery.

Keywords: *Artificial Intelligence, Machine Learning, Drug Discovery, Chemical Synthesis, Materials Science, Computational Chemistry.*

**Correspondence:shyamal.m68@gmail.com*

Introduction

Artificial Intelligence and Machine Learning are rapidly developing tools that have become the core of numerous fundamental and technical disciplines, among which chemistry is one of the most important and promising disciplines (Cun et al., 2015). The combination of these two approaches has stimulated abilities for unprecedented innovation. These methods can also be applied in data mining, molecular simulations, and property estimation, significantly influencing drug design, materials science, and chemical manufacturing (Goh et al., 2017). Thus, this introduction aims to familiarize the reader with the importance of AI and ML in revolutionizing chemistry, as well as their underlying concepts, methods, and applications (Schneider et al., 2018).

The history of AI can be traced back to the mid-twentieth century, with prominent figures such as Alan Turing and John McCarthy (Butler et al., 2018). However, significant progress in the field was not achieved until recent decades, driven by advancements in computational power, algorithms, and data availability (Coley et al., 2019). At the same time, the growth rate of the field of Machine Learning, an AI subfield accelerated. Concurrently, the rapid growth of machine learning, a subfield of AI, has been fueled by improvements in statistical modeling, optimization techniques, and neural networks (Segler et al., 2017).

Computational methods have been employed in chemistry since the advent of computers in the mid-twentieth century (Kuhn et al., 2013). Initially, these methods were primarily used to solve quantum mechanical problems to determine molecular structures and properties. Of late, the field of computational chemistry has broadened in line with the emergence of access to high performance computing and the implementation of Artificial Intelligence and Machine Learning (Lo et al., 2018). Today AI and ML are part of numerous aspects of chemical research, providing unparalleled opportunities for big data analysis as well as computer-aided molecular design and modeling (Coley et al., 2019).The primary applications of AI and ML in the field of chemistry are summarized in **Table 1** (Schneider et al., 2018; Noé et al., 2020; Butler et al., 2018; Schwaller et al., 2019; Korolev et al., 2020; Vamathevan et al., 2019; Wu et al., 2018; Mattei et al., 2019; Goh et al., 2017; Pfrommer et al., 2018). However, this review covers only the four main aspects (drug discovery, synthetic chemistry, material science, and computational chemistry) of AI and ML being used in chemistry.

Table 1: Use of AI and ML in multiple fields of chemistry.

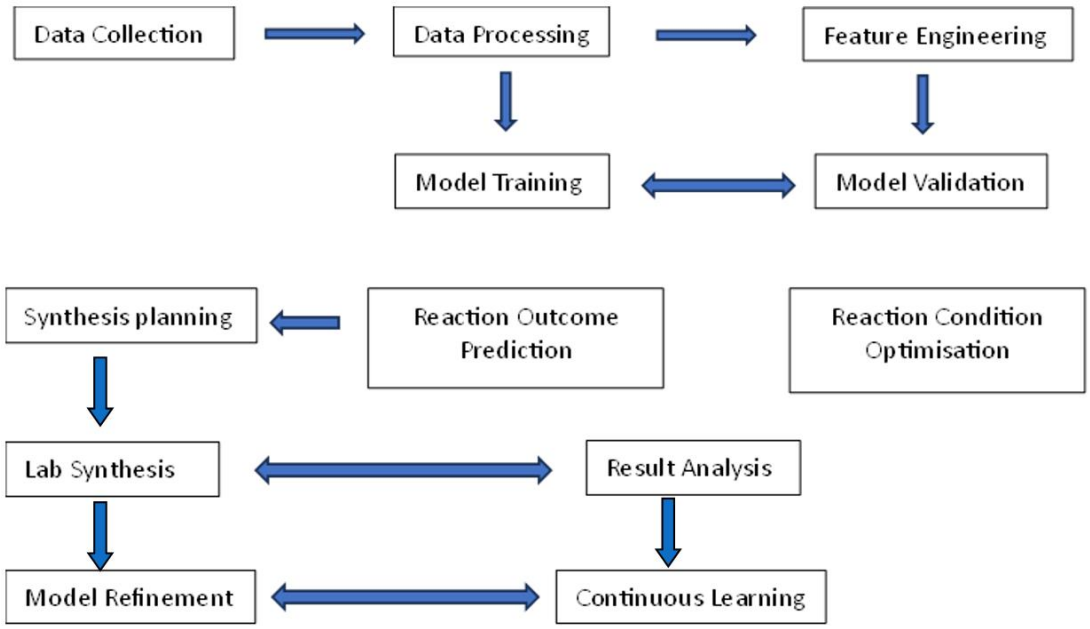
Application	Description	Examples
Drug Discovery	AI and ML models are used to predict the biological activity of compounds, identify potential drug candidates, and optimize their properties	Identifying new drug candidates, predicting drug-target interactions, and optimizing drug formulations.
Molecular Simulation	AI aids in simulating molecular dynamics, predicting molecular structures, and understanding complex chemical reactions	Simulating protein folding, predicting molecular behavior, and exploring reaction mechanisms.
Material Science	AI helps in discovering and designing new materials with desired properties by analyzing vast datasets and predicting material behavior.	Developing new catalysts, designing high-performance materials, and optimizing material properties.
Chemical Synthesis	ML models predict the outcomes of chemical reactions and suggest optimal synthesis pathways for target molecules.	Predicting reaction products, optimizing synthesis routes, and automating synthetic chemistry.
Spectroscopy Analysis	AI analyzes spectral data (e.g., NMR, IR, MS) to identify compounds and their structures, and to monitor reaction progress.	Automated interpretation of NMR spectra, identifying compounds from mass spectrometry data.

**Correspondence:shyamal.m68@gmail.com*

Cheminformatics	AI-driven analysis of chemical data (e.g., databases of compounds and reactions) to find patterns and make predictions.	Predicting chemical properties, finding similar compounds, and virtual screening of chemical libraries.
Toxicology and Safety	AI models predict the toxicity and environmental impact of chemical substances, helping to ensure safety and regulatory compliance.	Predicting adverse effects of chemicals, environmental impact assessment, and regulatory toxicology.
Formulation Science	AI optimizes the formulation of chemical products, such as pharmaceuticals, agrochemicals, and consumer goods.	Optimizing drug formulations, developing stable agrochemical mixtures, and improving product performance.
Analytical Chemistry	AI enhances the analysis and interpretation of complex data from analytical instruments, improving accuracy and efficiency.	Automated data analysis from chromatography, improving detection limits in analytical techniques.
Process Optimization	AI optimizes chemical processes in industrial settings, improving yield, efficiency, and sustainability.	Optimizing reaction conditions in manufacturing, reducing waste, and improving energy efficiency.

Fundamental Concepts and Methodologies

The core to the processes of AI and ML is the ability to learn from the data, to make decisions or predictions (Hansen et al., 2009). ML algorithms can be preliminarily distinguished according to the type of learning: supervised, unsupervised, or reinforcement learning with different learning goals and approaches (Rupp et al., 2012). Furthermore, reinforcement learning algorithms have been used to enhance reaction conditions, and to generate new molecules with desired properties through a process of trial and error in terms of chemical space (Coley et al., 2019). The methodology for chemical synthesis using AI & ML involves several steps that can be represented in a flow diagram **(Scheme 1)** (Schwaller et al., 2019).



Scheme 1. The common methodology of using AI & ML in different fields of chemistry.

*Correspondence:shyamal.m68@gmail.com

Applications in Drug Discovery

The implementation of ML approaches in drug discovery has fast-forwarded the discovery of novel therapeutic assets based on the structure-toxicity profile analysis of large chemical data sets (Yang et al., 2019). Pharmaceuticals remain one of the largest domains where AI and ML have applied their presence and disrupted the conventional process of drug discovery. Traditionally, drug discovery was a process prolonged in time and rather expensive, involving a high level of failures and growing costs. Still, AI and ML can save the time to get through several steps of the drug discovery process, including target identification, lead optimization, clinical trial design, and post-marketing surveillance (Schneider et al., 2018).

Virtual screening and quantizing are amongst the most active fields in the engagements of AI in the drug discovery. Tools such asAlphaFold (Jumper et al., 2021), DeepChem (Wu et al., 2018), AutoDock Vina (Trott et al., 2010), REINVENT (Blaschke et al., 2020), ADMETlab (Dong et al., 2018), MOSES (Polykovskiy et al., 2020) are used to learn models for the interaction of drugs and targets based on diverse datasets of chemical compounds and related biological activities (Lo et al., 2018). These models can then be used with success to rationally select compounds for experimental testing based on their predicted binding affinities and selectivity as well as pharmocokinetic properties.

In addition, AI, and ML help in the refinement of lead compounds by evaluating the structure-activity relationship (SAR) rationally as well as molecular designing. Using ML algorithms and mathematical equations, the large datasets of chemical structures along with the bioactivity profiles can be used to determine several critical molecular features. These mathematical approaches can help in continuous refinement through successive rounds of modification of biomolecules to improve their potency, selectivity, and safety (Jumper et al., 2021; Bombarelli et al., 2018; Ramsundar et al., 2019).

Molecules are represented in forms suitable for deep learning models:

1.Molecular Representation

i) SMILES Strings: Linear notation converted into embeddings for sequence models.

ii) Molecular Graphs:

Nodes (**v**) represent atoms.

Edges (**e**) represent bonds.

Graph Representation

A molecule as a graph

$$G = (V, E),$$

V: Set of atoms.

E: Set of bonds.

2. Feature Extraction with Graph Neural Networks (GNNs)

i) Message Passing Framework

For a node **v** in a graph **G** feature updates occur iteratively:

$$h_v^{(t+1)} = f\left(h_v^{(t)}, \{h_u^{(t)} : u \in \mathcal{N}(v)\}, e_{uv}\right)$$

$e_{uv} : Featuresofedge(u, v)$

$\mathcal{N}(v)$: Neighborhood of node (v).

$h_v^{(t)}$: Feature vector of node v at iteration t.

ii) Readout Function

Aggregates node features to represent the entire graph:

$$h_G = \text{READOUT}\left(\{h_v^{(T)} \middle| v \in G\}\right)$$

h_G is used for property prediction.

*Correspondence:shyamal.m68@gmail.com

3. Property Prediction (Regression/Classification)

i) Regression Model

Predict molecular properties like binding affinity (y):

$$\hat{y} = f(h_G; \theta)$$

where f is a neural network with parameters θ

ii) Classification Model

For binary outcomes (e.g., active/inactive), the output probability is:

$$\hat{y} = \sigma\big(f(h_G; \theta)\big)$$

where $\sigma(x) = \frac{1}{1+e^{-x}}$ is the sigmoid activation function.

iii) Loss Functions

Mean Squared Error (MSE) for regression:

$$\mathcal{L}\text{ME} = \frac{1}{N}\sum_{i=1}^{N}(\hat{y}_i - y_i)^2$$

Binary Cross-Entropy for classification:

$$\mathcal{L}\text{BE} = -\frac{1}{N}\sum_{i=1}^{N}[y_i \log(\hat{y}_i) + (1 - y_i)\log(1 - \hat{y}_i)]$$

4. Molecular Generation with Variational Autoencoders (VAEs)

VAEs generate novel molecules by learning a latent representation.

i) Encoding: Input molecule x is encoded into a latent vector z:

$$q_\phi(z|x) = \mathcal{N}\left(z; \mu_\phi(x), \sigma_\phi^2(x)\right)$$

where ϕ are encoder parameters.

ii) Decoding: The decoder generates a molecule from z:

$$p_\theta(x|z)$$

where θ are decoder parameters.

iii) Loss Function: Combines reconstruction and regularization:

$$\mathcal{L} = \mathcal{L}\text{rcn} + \beta\mathcal{L}_\text{K}$$

Reconstruction Loss:

$$\mathcal{L}\text{rcn} = -E_{q_\phi(z|x)}[\log p_\theta(x|z)]$$

KL Divergence Regularization:

$$\mathcal{L}_\text{K} = D_\text{KL}\big[q_\phi(z|x)|p(z)\big]$$

5. Drug-Target Interaction Prediction

i) Matrix Factorization

Predict drug-target binding affinity (y_{ij}) by factorizing drug and target embeddings:

$$\widehat{y_{iJ}} = u_i^{\top v_j}$$

u_i: Drug embedding.

V_j: Target embedding.

ii) Deep Learning Prediction

Using concatenated features:

$$\widehat{y_{iJ}} = f\big([u_i, v_j]; \theta\big)$$

where f is a deep neural network.

6. Optimization with Reinforcement Learning (RL)

R optimizes molecular properties using a reward function R.

i) Q-Learning

**Correspondence:shyamal.m68@gmail.com*

For state s (current molecule) and action a (modification):

$$Q(s,a) \leftarrow Q(s,a) + \alpha\left[R + \gamma \max_{a'} Q\,(s',a') - Q(s,a)\right]$$

α: Learning rate.
γ: Discount factor.
s′,a′: Next state and action.

ii) Reward Function

Defined for desired properties:

$$R = f(\text{binding affinity, toxicity}) - \text{penalty for undesired traits.}$$

Aside from the lead optimization, AI proximal entries significantly contribute to toxicity prognosis and adverse event characterization, allowing the early assessment of safety-related issues and furthermore, the selection of the candidates during preclinical and clinical stages (Segler et al., 2017). In this way, ML algorithms are capable of learning chemical structure and toxicological properties from multitude sources of chemical, biological, and clinical data, to create a reasonable risk profile to guide the evaluation of risk and possible protective measures.

The exploration of the AI and ML in drug discovery is the key enhancer of the efficient identification and development of brand-new safe and efficacious therapeutics, as well as the minimization of time and costs necessary for trial-and-error approaches.

Applications in Synthetic Chemistry

In the domain of chemical synthesis, AI and ML offer unprecedented opportunities to accelerate discovery, optimize processes, and reduce costs. By leveraging advanced algorithms, and different tools like ASKCOS (Grzybowski et al., 2018), IBM RXN (Schwaller et al., 2018), AiZynthFinder (Genheden et al., 2020), ChemOS (Roch et al., 2020) researchers are now able to navigate the complex chemical space more efficiently.

i) Optimization of Reaction Conditions

Machine learning models can analyse vast datasets of reaction outcomes to predict optimal conditions such as temperature, pressure, catalysts, and solvent choices; for instance, gradient boosting algorithms or neural networks are trained on experimental data to fine-tune these parameters, thus reducing the need for exhaustive experimentation (Guzik et al., 2018; Cova et al., 2020).

Machine learning predicts rate constants (k) using Arrhenius-type relationships:

$$k = A \cdot e^{-\frac{E_a}{RT}}$$

where A is the pre-exponential factor, Ea is the activation energy, R is the gas constant, and T is the temperature. E_a and A can be predicted using ML models.

ii) Predicting Reaction Outcomes

When fed the information about molecular structures and the reaction conditions, the ML models will predict the conversion, selectivity, and potential side reactions of a chemical process (Schwaller et al., 2019). This predictive capability is particularly beneficial in compound multistep syntheses because byproducts may interfere with the synthesis procedure. Bayesian networks and deep learning-based models have been predicted to enhance the reaction predictability as well (Coley et al., 2018; Gao et al., 2020).

Chemical reaction prediction can be modelled as:

$$P(y|X) = f(X,\Theta)$$

where P(y|X) is the probability of a reaction outcome y, given the input features X (e.g., reactants, reagents, and conditions), and f is a predictive model parameterised by Θ.

*Correspondence:shyamal.m68@gmail.com

iii) Designing Synthesis Pathways

The choice of synthetic strategy for the target molecules is one of the most important components in the design of a chemical synthesis. Previously this has been done qualitatively by chemists using their experience and estimations together with the concept of retrosynthetic analysis. This is where AI platforms including the retrosynthesis planning tools help in suggesting the probable pathways of synthesis just by considering the structure of a certain molecule (Segler et al., 2018). These tools leverage pre-existing specifications concerning reactions and use random strategies such as the Monte Carlo tree search procedure or the graph neural network to detect suitable paths (Yang et al., 2020).

Retrosynthesis can be framed as a tree-search optimization:

$$\text{Score}(S) = \sum_{i=1}^{n} g(S_i) + \lambda \cdot h(S)$$

where g(S_i) evaluates the likelihood of sub-reactionsSi; h(S)is a heuristic function estimating the cost of synthesis paths, and λis a weighting parameter.

iv) Exploration of Chemical Space

It is impossible to conceive how vast the chemical space (the space of all possible molecules and their permutations) is. Machine learning and artificial intelligence allow for this idea through optimizing the computation in high dimensions and using a system that can catch what would not necessarily be naturally recognized by human instinct alone (Bombarelli et al., 2018). Among the generational models which propose new molecules with specific properties, VAE and GAN are the most suitable (Elton et al., 2019; Lengeling et al., 2018).

Using deep generative models:

$$z = E(x), \quad x' = D(z)$$

where E is the encoder that maps a molecule x to a latent vector z, and D is the decoder generating new molecules x′from z.

v) Catalyst Design and Development

A promising method for the synthesis of new catalysts is often associated with exhaustive experimental testing. Machine learning improves this process much faster as it focuses on the correlation between the structure of the catalyst and its performance (Nandy et al., 2020). Through the data analysis of results obtained by computational simulations and experiments, ML models can propose possible catalyst candidates, which will essentially shorten the period and expenses needed in the catalyst design (Butler et al., 2018; Toyao et al., 2020).

Catalyst activity (AAA) can be correlated with ML-predicted descriptors:

$$A = f(\{X_i\}) = w_1X_1 + w_2X_2 + \cdots + w_nX_n + b$$

where {Xi} are features such as electronic properties or coordination numbers.

vi) High-throughput Experimentation

Molecular biology: high through-put experimentation and training of AI robots and autoionic have increased the rate of data production and analysis (MacLeod et al., 2020). These systems use machine learning to optimize the experiment design on the fly, based on feedback (Jensen et al., 2019). As an example, self-formed synthesis machines can carry out independent reaction, assay results, and adjust parameters repeatedly, which effectively improves the processing capabilities of research channels (Huo et al., 2017).

Screening molecular candidates in a large chemical space:

$$\text{Similarity}(x_1, x_2) = \frac{\text{dot}(x_1, x_2)}{|x_1||x_2|}$$

where $x_1$and x_2 are feature vectors for molecules, and similarity guides the screening process.

**Correspondence:shyamal.m68@gmail.com*

vii) Process Optimization and Scale-Up

Advanced algorithms provide support in fine-tuning of process parameters for the reproducibility and scalability purposes (Ahneman et al., 2018). Thanks to the data obtained at the pilot scale, these models reveal critical constraints and possible challenges in real-world industrial environments and thus, ease the move towards commercial scale production (Toyao et al., 2018; Schwaller et al., 2020).

Yield prediction often uses probabilistic models:

$$Y = \sigma(w^T X + b)$$

where Y is the yield, X are molecular and reaction descriptors, and σ is a sigmoid activation function.

viii) Sustainability and Green Chemistry

The use of AI and ML for chemical synthesis also has tendencies to be in harmony with the principles of green chemistry (1, 2, and 6) since the generated information does not include generation of waste and the utilization of poisonous reagents. Such models allow chemists to spot the reactivity that employs renewable raw materials, less hazardous solvents, and energy-friendly conditions (Li et al., 2020) which, in return, enhances the formation of environmentally friendly processes as far as the economy is considered (Young et al., 2017; Coley et al., 2019).

Optimizing reaction conditions for sustainability:

$$\text{minimize: EI} = \frac{M_{\text{waste}}}{M_{\text{product}}}$$

where EI is the environmental impact factor, predicted using ML for various conditions.

Applications in Materials Science

The use of artificial intelligence in the design of materials has led to enhanced creation of new materials with specific characteristics for the specific need such as energy storage and catalysis Gabriel et al., 2019). In material science, AI and ML tools like MatGAN (Ren et al., 2018), Materials Project (Jain et al., 2013), CGCNN (Xie et al., 2018), MEGNet (Chen et al., 2019) have been identified as state-of-the-art technologies for enhancing the rate of development and enhancement of new qualities of material with diverse potential uses such as energy conversion and storage, catalytic activity, and sensing. Conventional methods of materials discovery points back to hunch, rule of thumb, and finger-crossing and so, entails long and resultant cycle performances along with minimal mapping of the chemical space (Butler et al., 2018). Given below is a general mathematical framework for these applications (Scarselli et al., 2009; Duvenaud et al., 2015; Gilmer et al., 2017; Xu et al., 2019; Kingma et al., 2014; Higgins et al., 2017; LeCun et al., 2015; Goodfellow et al., 2016).

1. Representation of Materials

Materials are represented as structured data (graphs, tensors, or descriptors):

Graph Representation: Materials are treated as graphs G=(V,E) where:

V represents nodes (atoms or molecular units).

E represents edges (bonds or interactions).

Mathematical representation of material graphs:

$$G_i = (V_i, E_i), \quad V_i \in R^d, \quad E_i \in R^e$$

where ddd and eee are the feature dimensions for nodes and edges.

2. Feature Encoding

Deep learning models encode the features of a material using embedding techniques:

Atomic Embeddings:

$$h_i^0 = f_{\text{embed}}(x_i), \quad x_i \in V$$

**Correspondence:shyamal.m68@gmail.com*

where h_i^0 is the initial feature vector for node iii, and f_{embed} is an embedding function (e.g., one-hot encoding or learned embeddings).

Bond Embeddings:

$$h_{ij} = g_{\text{embed}}(e_{ij}), \quad e_{ij} \in E$$

3. Graph Neural Networks (GNNs) for Material Properties

GNNs aggregate information from neighbouring nodes and edges to learn material properties:

Node feature updates:

$$h_i^{(l+1)} = \sigma\left(W^{(l)} \cdot \text{AGG}\left(\{h_i^{(l)}, h_j^{(l)}, h_{ij} : j \in \mathcal{N}(i)\}\right)\right)$$

where:

$h_i^{(l+1)}$ is the updated feature vector for node iii at layer l+1l+1l+1.

N(i) is the set of neighbours for node iii.

$W^{(l)}$ are learnable weight matrices.

AGGis an aggregation function (e.g., summation, averaging).

σ is a non-linear activation function (e.g., ReLU).

Global material property prediction:

$$\hat{y} = f_{\text{readout}}\left(\text{READOUT}\left(\{h_i^{(L)}\}_{i=1}^{n}\right)\right)$$

where READOUT aggregates node-level information into a graph-level representation, and $f_{readout}$ is a predictor (e.g., fully connected network).

4. Variational Autoencoders (VAEs) for Material Design

VAEs learn a latent representation of materials and generate new structures:

Latent Space Encoding:

$$z = \mu + \epsilon \cdot \exp(\log(\sigma^2)), \quad \epsilon \sim \mathcal{N}(0,1)$$

where:μ and σ are the mean and variance of the latent distribution.

z is the latent vector for a material.

Reconstruction Loss:

$$\mathcal{L}\text{rcn} = |X - X'|^2$$

where X and X′ are the original and reconstructed material features.

Regularization Loss (Kullback-Leibler Divergence):

$$\mathcal{L}_{\text{K}} = D_{\text{KL}}\left(q_\phi(z|X)|p(z)\right)$$

Total Loss:

$$\mathcal{L} = \mathcal{L}\text{rcn} + \beta \cdot \mathcal{L}_{\text{K}}$$

5. Predicting Material Properties

Regression or classification models predict material properties:

Output Prediction:

$$\hat{y} = f_\theta(X), \quad f_\theta = \text{Neural Network Model}$$

Where $\hat{y}$ is the predicted property (e.g., band gap, conductivity), and f_θ is the trained model.

Loss Function:

For regression: $\mathcal{L} = \frac{1}{N}\sum_{i=1}^{N}(\hat{y}_i - y_i)^2$

For classification: $\mathcal{L} = -\frac{1}{N}\sum_{i=1}^{N}(y_i \log(\hat{y}_i) + (1 - y_i)\log(1 - \hat{y}_i))$

6. Implementation and Applications

Crystal Structure Prediction: ML models predict stable crystal structures based on atomic composition and features.

Example: Predicting lattice parameters using GNNs.

*Correspondence:shyamal.m68@gmail.com

Property Optimization: Bayesian optimization identifies optimal material candidates by combining ML-predicted properties with experimental constraints.
Generative Design: VAEs and GANs create novel molecular frameworks or crystalline materials tailored to desired properties.

Instead, AI and ML are high-data ways of not only materials design and optimization where big data of experimental and computational data are used to find hidden patterns and structure-property relationship and make better decisions. For example, using real features that mark material properties, in other words, chemical content, crystallography, and processing parameters, it is possible to train the ML algorithms for predictive models that can quickly and effectively screen a range of material candidates with required and useful properties. These various applications of AI in materials science include; one of the most important applications being the design of novel battery materials with high energy density, high cyclability, and safe. Using the large amount of experimental and computational data regarding the electrochemical properties of battery materials, the ML algorithms can select the appropriate types of electrode materials, electrolytes, and nanoscale structures, which can improve the battery energy output and cycling stability (Xie & Grossman, 2018; Schütt et al., 2018; Ramsundar et al., 2019).

In general, the application of AI and ML in materials sciences is full of opportunities to enhance the discovery and performance of new materials as well as to advance the conveyance of scientific ideas into technologies across many fields of technology.

Applications in Computational Chemistry

Computational chemistry is the science that strives to comprehend chemical systems by means of theoretical simulation. This discipline has been greatly transformed by the development of AI as well as ML tools like SchNet (Schütt et al., 2018), DeepChem (Wu et al., 2018), RDKit (Landrum, 2006), Gaussian, PySCF (Sun et al., 2018), Chempro (Goh et al., 2017). Such methods provide fresh approaches for enhancing the rate of determination of molecular properties, for providing insights into the detailed process of chemical reactions and for providing design principles for chemical systems (Kipf et al., 2017; Gilmer et al., 2017; Schütt et al., 2017; Behler et al., 2007; Brockherde et al., 2017; Kingma et al., 2014; Schwaller et al., 2019; Caruana, 1997; Vaswani et al., 2017).

1. Molecular Graph Representation

Molecules are represented as graphs:

$$G = (V, E), \quad V = \{v_1, v_2, \ldots, v_n\}, \quad E = \{e_{ij} \mid i, j \in V\}$$

where V is the set of atoms (nodes) and E is the set of bonds (edges).

2. Atomic and Bond Feature Initialization

$$h_i^0 = f_{\text{embed}}(x_i), \quad x_i \in R^{d_v}, \quad e_{ij}^0 = g_{\text{embed}}(b_{ij}), \quad b_{ij} \in R^{d_e}$$

Here, $x_i$represents atomic features, b_{ij} represents bond features, and f_{embed},$g_{\text{embed}}$are embedding functions.

3. Graph Neural Network (GNN) for Message Passing

Message passing updates node features iteratively:

$$m_{ij}^{(l)} = \phi_m\left(h_i^{(l)}, h_j^{(l)}, e_{ij}\right)$$

$$h_i^{(l+1)} = \phi_h\left(h_i^{(l)}, \sum_{j \in \mathcal{N}(i)} m_{ij}^{(l)}\right)$$

*Correspondence:shyamal.m68@gmail.com

where $m_{ij}^{(l)}$ is the message between atoms i and j, $\mathcal{N}(i)$is the set of neighbours of i, andϕ_m, ϕ_h are differentiable functions (e.g., neural networks).

4. Molecular Property Prediction

A graph-level representation is computed using a readout function:

$$h_G = \text{READOUT}\left(\{h_i^{(L)}\}_{i=1}^{n}\right)$$

$$\hat{y} = f_{\text{NN}}(h_G)$$

where READOUTcan be summation, averaging, or attention mechanisms, and f_{NN} is a neural network predicting molecular properties like energy, dipole moment, or HOMO-LUMO gaps.

5. Quantum Chemistry Prediction (Schrödinger Equation)

In quantum chemistry, deep learning approximates solutions to the Schrödinger equation:

$$\hat{H}\Psi = E\Psi$$

where $\hat{H}$ is the Hamiltonian operator, E is the molecular energy, and Ψ is the wavefunction. Neural networks approximate Edirectly:

$$E \approx f_\theta(X)$$

where X is the molecular descriptor (e.g., Coulomb matrix or density grid), and f_θ is a neural network with parameters θ.

6. Potential Energy Surface (PES) Learning

The PES, which relates molecular structure to energy, is modeled as:

$$E(R) \approx f_\theta(R)$$

where R is the 3D atomic coordinate matrix, and f_θ is a neural network trained on ab initio data.

7. Loss Functions

a) Property Prediction Loss:

$$\mathcal{L}_{\text{property}} = \frac{1}{N}\sum_{i=1}^{N}(\hat{y}_i - y_i)^2$$

where $\hat{y}_i$ is the predicted property and y_i is the ground truth.

b) Wavefunction Loss (if approximating Ψ):

$$\mathcal{L}_{\text{wavefunction}} = |\Psi - \hat{\Psi}|^2$$

8. Transfer Learning for Quantum Properties

Pre-trained models on quantum datasets (e.g., QM9) are fine-tuned:

$$\mathcal{L} = \mathcal{L}_{\text{property}} + \lambda|\theta - \theta_0|^2$$

where θ_0 *is* the pre-trained weights and λ controlthe regularization term to ensure the learned weights θ do not deviate significantly from the pre-trained model parameters.

9. Neural Network Potentials (NNP)

Neural network potentials (like ANI or SchNet) are used to predict molecular forces and energies:

$$E_{\text{total}} = \sum_{i=1}^{N} f_{\text{atom}}(x_i)$$

where f_{atom} predicts the contribution of atom i based on its local environment.

The gradient of the energy gives forces:

$$F_i = -\frac{\partial E_{\text{total}}}{\partial R_i}$$

where $R_i$are the atomic coordinates.

*Correspondence:shyamal.m68@gmail.com

10. Electron Density Prediction

Deep learning models predict the electron density $\hat{\rho}(r)$ directly:

$$\hat{\rho}(r) = f_{\text{NN}}(r, Z, R)$$

where r is the spatial position, Z is the atomic numbers, and R is the atomic coordinates.

11. Orbital Energy Prediction (HOMO/LUMO)

Deep learning models can predict orbital energies:

$$E_{\text{HOMO}}, E_{\text{LUMO}} \approx f_{\theta}(C)$$

where C represents molecular descriptors such as molecular fingerprints or Coulomb matrices.

12. Reaction Prediction

Deep learning for chemical reactions can be expressed as:

$$P(\text{products} \mid \text{reactants}) = \text{softmax}\big(f_{\theta}(\text{reactants})\big)$$

where f_{θ} is a neural network that scores potential products, and softmax normalizes the scores into probabilities.

13. Multi-task Learning for Computational Chemistry

A single neural network can predict multiple properties simultaneously:

$$\mathcal{L} = \sum_{k=1}^{K} \alpha_k \mathcal{L}_k$$

where K is the number of tasks, $\mathcal{L}_k$ is the loss for the K-th task, and α_k are task-specific weights.

14. Attention Mechanisms for Atomic Interactions

Attention is often used to weigh atomic interactions:

$$\alpha_{ij} = \frac{\exp(e_{ij})}{\sum_{k \in \mathcal{N}(i)} \exp(e_{ik})}, \quad e_{ij} = f_{\text{score}}(h_i, h_j)$$

where α_{ij} is the attention score between atom i and j, and f_{score} is a neural network.

15. Generative Models for Molecular Design

Variational autoencoders (VAEs) or generative adversarial networks (GANs) are used to generate novel molecules:

a) VAE Latent Space: $z = \mu + \sigma \cdot \epsilon, \quad \epsilon \sim \mathcal{N}(0,1)$

b) Molecule Reconstruction: $\hat{X} = f_{\text{decode}}(z)$

These equations represent the foundational methods and concepts in applying AI and ML to computational chemistry, covering property prediction, quantum mechanics, reaction dynamics, and molecular design.

Another important area of computational chemistry benefits from the integration of artificial intelligence is the enhancement of computation time of quantum mechanical calculations by the creation of data-driven models.[94] It is to be noted that although quantum mechanics-based calculations are highly reliable in most of the cases, the time required for computation is relatively very high and hence makes it difficult if not impossible to screen large chemical systems. These surrogate models are trained using neural networks with architectures that mimic the general operation of quantum chemistry programs on quantum mechanical calculation input and output data sets. The use of these cheap surrogate models supports rapid chemoinformatics and high throughput virtual screening (Ramakrishnan et al., 2015).

Furthermore, there are the approaches to predict the molecular properties and reactivity patterns based on the empirical models built on the core sets of experimentally measured values. For instance, using datasets, containing chemical structures with biological activities, ML algorithms can learn QSAR models and predict pharmacological profiles, toxicological endpoints, and the

*Correspondence:shyamal.m68@gmail.com

environmental fate. Likewise, there are quantitative structure property relationship (QSPR) models which provide prediction of thermodynamic, electronic and transport properties of the chemical compounds based on the molecular descriptor and physical property datasets (Rupp et al., 2012).

In addition, using AI and ML methods, one can study reaction mechanisms and predict the outcome of the reaction based on data. Just as reaction databases and computational simulations, ML algorithms can determine reaction templates, estimate both bond-breaking and forming events and offer the possible reaction pathways for further experimental optimization (Schütt et al., 2018; Gilmer et al., 2017; Rupp et al., 2012; Jørgensen et al., 2018; Gilmer et al., 2017). These can also foresee regio selectivity, stereo selectivity and chemoselectivity of transformations given patterns from training samples which can be very useful in considered synthesis and retrosynthetic analysis.

In the areas of molecular modeling and molecular property and function prediction, AI and ML can leapfrog current challenges and give insights into chemical systems, and in the design of new molecular structures and functions, they can enhance and inspire new concepts and design strategies.

Challenges and Future Directions

While the integration of AI and ML in chemistry holds immense promise for advancing research and innovation, several challenges must be addressed to fully realize their potential and ensure responsible deployment in practice. One of the key challenges pertains to the interpretability and explainability of ML models, especially in complex chemical systems where underlying relationships may be non-linear and multifaceted. Ensuring transparency and robustness of AI-driven predictions is essential for building trust and confidence in the reliability of predictive models, particularly in safety-critical applications such as drug discovery and materials design.

Moreover, the availability and quality of data pose significant challenges for training ML models, as chemical datasets are often heterogeneous, incomplete, and biased. Addressing data scarcity and data quality issues requires concerted efforts to curate, standardize, and share chemical data repositories, thereby enabling broader access.

Conclusions

In conclusion, the integration of Artificial Intelligence (AI) and Machine Learning (ML) in chemistry represents a transformative paradigm shift, offering unprecedented opportunities for accelerating research and innovation across diverse domains. From drug discovery and materials science to computational chemistry and chemical synthesis, AI and ML have demonstrated remarkable capabilities in predictive modeling, data analysis, and decision-making, enabling researchers to tackle complex challenges and discover new molecules and materials with tailored properties and functionalities. Despite the immense progress achieved, several challenges remain, including the interpretability of ML models, data quality issues, and ethical considerations. Moving forward, interdisciplinary collaboration, data sharing initiatives, and advances in algorithmic transparency will be crucial for harnessing the full potential of AI and ML in chemistry while ensuring responsible innovation and societal benefit. By leveraging the synergistic interplay between human expertise and machine intelligence, we can unlock new frontiers in chemical research and pave the way for transformative discoveries with profound implications for healthcare, energy, and sustainability.

Acknowledgement

SM extends gratitude to the Principal of New Alipore College, Kolkata, for his unwavering encouragement and invaluable support at every step. His guidance has been instrumental, and without his assistance, the success of this review would not have been possible.

**Correspondence:shyamal.m68@gmail.com*

References

Ahneman, D. T., Estrada, J. G., Lin, S., Dreher, S. D., & Doyle, A. G. (2018). Predicting Reaction Performance in C–N Cross-Coupling Using Machine Learning. *Science*, 360(6385), 186–190. DOI: 10.1126/science.aar5169.

Aspuru-Guzik, A., & Persson, K. A. (2018). Materials Acceleration Platform: Accelerating Advanced Energy Materials Discovery by Integrating High-Throughput Methods and Artificial Intelligence. *Chemistry of Materials*, 30(10), 3533–3543. DOI: 10.1021/acs.chemmater.8b00183.

Behler, J., & Parrinello, M. (2007). *Generalized neural-network representation of high-dimensional potential-energy surfaces*. Physical Review Letters, 98(14), 146401. https://doi.org/10.1103/PhysRevLett.98.146401

Blaschke, T., De Fabritiis, M., Laino, T., & Reymond, J.-L. (2020). REINVENT 2.0: An AI tool for de novo drug design. *Journal of Chemical Information and Modeling*, 60(10), 5918–5922.

Brockherde, F., Vogt, L., Li, L., et al. (2017). *Bypassing the Kohn-Sham equations with machine learning*. Nature Communications, 8(1), 872. https://doi.org/10.1038/s41467-017-00839-3.

Butler, K. T., Davies, D. W., Cartwright, H., Isayev, O., & Walsh, A. (2018). Machine learning for molecular and materials science. Nature, 559(7715), 547–555. DOI: 10.1038/s41586-018-0337-2.

Caruana, R. (1997). *Multitask learning*. Machine Learning, 28(1), 41–75. https://doi.org/10.1023/A:1007379606734.

Chen, C., et al. (2019). "Graph Networks as a Universal Machine Learning Framework for Molecules and Crystals." *Chemistry of Materials*, 31(9), 3564–3572.

Coley, C. W., Eyke, N. S., Jensen, K. F., & Mitragotri, S. (2019). Artificial intelligence for chemical discovery. Nature Reviews Chemistry, 1(3), 0138. DOI: 10.1038/s41570-019-0138-1.

Coley, C. W., Green, W. H., & Jensen, K. F. (2018). Machine Learning in Computer-Aided Synthesis Planning. *Accounts of Chemical Research*, 51(5), 1281–1289. DOI: 10.1021/acs.accounts.8b00087.

Coley, C. W., Jin, W., Rogers, L., Green, W. H., & Jensen, K. F. (2019). A Graph-Convolutional Neural Network Model for the Prediction of Chemical Reactivity. *Chemical Science*, 10(2), 370–377. DOI: 10.1039/C8SC04228D.

Coley, C. W., Thomas, D. A., Lummiss, J. A., Jaworski, J. N., Breen, C. P., Schultz, V., & Hart, T. (2019). A robotic platform for flow synthesis of organic compounds informed by AI planning. Science, 365(6453), eaax1566. DOI: 10.1126/science.aax1566.

Coley, C. W., Thomas, D. A., Lummiss, J. A., Jaworski, J. N., Breen, C. P., Schultz, V., & Hart, T. (2019). A robotic platform for flow synthesis of organic compounds informed by AI planning. Science, 365(6453), eaax1566. DOI: 10.1126/science.aax1566.

Cova, T. F., & Pais, A. A. C. C. (2019). Deep Learning for Deep Chemistry: Optimizing the Prediction of Chemical Reactions. *Chemical Science*, 10(35), 8187–8201. DOI: 10.1039/C9SC02766K.

Dong, J., Wang, D., Zhang, S., Wang, L., Luo, Y., & Wang, T. (2018). ADMETlab: A platform for systematic ADMET evaluation based on a comprehensive database. *Journal of Cheminformatics*, 10(1), 29.

Duvenaud, D. K., Maclaurin, D., Aguilera-Iparraguirre, J., et al. (2015). *Convolutional networks on graphs for learning molecular fingerprints*. Advances in Neural Information Processing Systems, 28, 2224–2232.

Elton, D. C., Boukouvalas, Z., Fuge, M. D., & Chung, P. W. (2019). Deep Learning for Molecular Design—a Review of the State of the Art. *Molecular Systems Design & Engineering*, 4(4), 828–849. DOI: 10.1039/C9ME00039A.

**Correspondence:shyamal.m68@gmail.com*

Gabriel R Schleder, Antonio C M Padilha, Carlos Mera Acosta, Marcio Costa, Adalberto Fazzio. "From DFT to machine learning: recent approaches to materials science–a review", Journal of Physics: Materials, 2019.

Gao, H., & Coley, C. W. (2020). The Synthesizability of Molecules Proposed by Generative Models. *Journal of Chemical Information and Modeling*, 60(12), 5714–5723. DOI: 10.1021/acs.jcim.0c00554.

Genheden, S., Thakkar, A., Chadimova, V., Reymond, J.-L., & Engkvist, O. (2020). AiZynthFinder: A fast, robust, and flexible open-source software for retrosynthetic planning. *Journal of Cheminformatics*, 12(1), 70.

Gilmer, J., Schoenholz, S. S., Riley, P. F., Vinyals, O., & Dahl, G. E. (2017). *Neural message passing for quantum chemistry*. Proceedings of the 34th International Conference on Machine Learning (ICML), 1263–1272. https://arxiv.org/abs/1704.01212

Goh, G. B., Hodas, N. O., & Vishnu, A. (2017). Deep learning for computational chemistry. Journal of Computational Chemistry, 38(16), 1291–1307. DOI: 10.1002/jcc.24764.

Gómez-Bombarelli, R., Wei, J. N., Duvenaud, D., Hernández-Lobato, J. M., Sánchez-Lengeling, B., Sheberla, D., & Aguilera-Iparraguirre, J., et al. (2018). Automatic chemical design using VAEs. *ACS Central Science*, 4(2), 268–276.

Goodfellow, I., Bengio, Y., & Courville, A. (2016). *Deep learning*. MIT Press.

Grzybowski, B. A., Szymkuc, R. G., Wlodarczyk-Pruszyński, J., Beker, K., & Kowalik, P. (2018). The Chematica software for automated retrosynthetic planning. *Nature Chemistry*, 10(5), 409–418.

Hansen, K., Mika, S., Schroeter, T., Sutter, A., & Ter Laak, A. (2009). Benchmark Data Set for in Silico Prediction of Ames Mutagenicity. Journal of Chemical Information and Modeling, 49(9), 2077–2081. DOI: 10.1021/ci900161g.

Higgins, I., Matthey, L., Pal, A., et al. (2017). *beta-VAE: Learning basic visual concepts with a constrained variational framework*. International Conference on Learning Representations (ICLR). https://openreview.net/forum?id=Sy2fzU9gl

Huo, H., & Rupp, M. (2017). Unified Representation of Molecules and Crystals for Machine Learning. *Machine Learning: Science and Technology*, 1(1), 1–13. DOI: 10.1088/2632-2153/ab6eb9.

Jain, A., et al. (2013). "The Materials Project: A materials genome approach to accelerating materials innovation." *APL Materials*, 1(1), 011002. https://materialsproject.org

Jensen, Z., Kim, E., Kwon, S., &Aspuru-Guzik, A. (2019). A Machine Learning Approach to Predicting Chemical Reaction Pathways. *Nature Reviews Chemistry*, 3(6), 347–360. DOI: 10.1038/s41570-019-0111-1.

Jørgensen, P. B., Petersen, A., Bligaard, T., & Thygesen, K. S. (2018). Machine learning-based screening of complex molecules for high-performance organic solar cells. *The Journal of Physical Chemistry Letters*, 9(5), 1055–1061.

Jumper, J., Evans, R., Pritzel, A., Green, T., Figurnov, M., Ronneberger, O., &Tunyasuvunakool, K., et al. (2021). Highly accurate protein structure prediction with AlphaFold. *Nature*, 596(7873), 583–589.

Kingma, D. P., & Welling, M. (2014). *Auto-encoding variational Bayes*. International Conference on Learning Representations (ICLR). https://arxiv.org/abs/1312.6114

Kingma, D. P., & Welling, M. (2014). *Auto-encoding variational Bayes*. ICLR. https://arxiv.org/abs/1312.6114

Kipf, T. N., & Welling, M. (2017). *Semi-supervised classification with graph convolutional networks*. ICLR. https://arxiv.org/abs/1609.02907

Korolev, V. V., Mitrofanov, A. Y., & Perminov, P. A. (2020). Artificial intelligence in the analysis of complex spectral data. *Journal of Analytical Chemistry*, 75(10), 1331-1340.

**Correspondence:shyamal.m68@gmail.com*

Kuhn, M., & Johnson, K. (2013). Applied predictive modeling. Springer Science & Business Media. DOI: 10.1007/978-1-4614-6849-3.

Landrum, G. (2006). "RDKit: Open-source cheminformatics." https://www.rdkit.org

Le Cun, Y., Bengio, Y., & Hinton, G. (2015). Deep learning. Nature, 521(7553), 436–444. DOI: 10.1038/nature14539.

Li, Z., Zhang, C., Feng, T., Ma, W., Liu, X., & Chen, J. (2020). Toward Sustainability: AI-Guided Green Chemistry. *Environmental Science & Technology Letters*, 7(8), 556–566. DOI: 10.1021/acs.estlett.0c00342.

Lo, Y. C., Rensi, S. E., Torng, W., Altman, R. B., & Liu, T. (2018). Machine learning in chemoinformatics and drug discovery. Drug Discovery Today, 23(8), 1538–1546. DOI: 10.1016/j.drudis.2018.05.009.

MacLeod, B. P., Parlane, F. G., Morrissey, T. D., Hase, F., Roch, L. M., Satchwell, M. L., Zaremba, O. T., Hein, J. E., & Cooper, G. J. (2020). Self-Driving Laboratory for Accelerated Discovery of Thin-Film Materials. *Science Advances*, 6(20), eaaz8867. DOI: 10.1126/sciadv.aaz8867.

Mattei, C., Pavia, G., & Trotta, M. (2019). Application of artificial intelligence in formulation development. *International Journal of Pharmaceutics*, 569, 118602.

Nandy, A., Duan, C., Kulik, H. J., & Hart, G. L. (2020). Computational Discovery of Transition-Metal Complexes with Crystal Field Theory and Machine Learning. *Chemistry of Materials*, 32(16), 7027–7037. DOI: 10.1021/acs.chemmater.0c02029.

Noé, F., Tkatchenko, A., Müller, K.-R., & Clementi, C. (2020). Machine learning for molecular simulation. *Annual Review of Physical Chemistry*, 71, 361-390.

Pfrommer, J., Gärtner, M., D'Angelo, L., Leonhard, K., & Henning, P. (2018). Optimizing chemical processes using machine learning: Recent advances and perspectives. *Journal of Chemical Information and Modeling*, 58(8), 1550-1559.

Polykovskiy, D., Zhebrak, M., Vetrov, A., Ivanenkov, A., Aladinskiy, V., Aliper, A., &Zhavoronkov, A., et al. (2020). MOSES: A benchmarking platform for molecular generation models. *Frontiers in Pharmacology*, 11, 565644.

Ramakrishnan, R., Dral, P. O., Rupp, M., & von Lilienfeld, O. A. (2015). Quantum chemistry structures and properties of 134 kilo molecules. Scientific Data, 1, 140022. DOI: 10.1038/sdata.2014.22.

Ramsundar, B., Eastman, P., Walters, P., & Pande, V. (2019). *Deep Learning for the Life Sciences*. O'Reilly Media.

Ren, F., et al. (2018). "Accelerated discovery of metallic glasses through iteration of machine learning and high-throughput experiments." *Science Advances*, 4(4), eaaq1566.

Roch, L. M., Häse, F., Kreisbeck, C., Tamayo-Mendoza, T., &Aspuru-Guzik, A. (2020). ChemOS: Orchestrating autonomous experimentation. *Science Advances*, 6(28), eabc2986.

Rupp, M., Tkatchenko, A., Müller, K.-R., & von Lilienfeld, O. A. (2012). Fast and accurate modeling of molecular atomization energies with machine learning. Physical Review Letters, 108(5), 058301. DOI: 10.1103/PhysRevLett.108.058301.

Sanchez-Lengeling, B., &Aspuru-Guzik, A. (2018). Inverse Molecular Design Using Machine Learning: Generative Models for Matter Engineering. *Science*, 361(6400), 360–365. DOI: 10.1126/science.aat2663.

Scarselli, F., Gori, M., Tsoi, A. C., Hagenbuchner, M., & Monfardini, G. (2009). *The graph neural network model*. IEEE Transactions on Neural Networks, 20(1), 61–80. https://doi.org/10.1109/TNN.2008.2005605

Schneider, G. (2018). Automating drug discovery. Nature Reviews Drug Discovery, 17(2), 97–113. DOI: 10.1038/nrd.2017.232.

Schütt, K. T., et al. (2018). "SchNet: A continuous-filter convolutional neural network for modeling quantum interactions." *The Journal of Chemical Physics*, 148(24), 241722.

**Correspondence:shyamal.m68@gmail.com*

Schütt, K. T., Kindermans, P. J., Sauceda, H. E., Chmiela, S., Tkatchenko, A., & Müller, K.-R. (2018). SchNet: A continuous-filter convolutional neural network for modeling quantum interactions. *Journal of Chemical Theory and Computation*, 14(11), 5820–5831.

Schütt, K., Kindermans, P., Sauceda, H. E., et al. (2017). *SchNet: A continuous-filter convolutional neural network for modeling quantum interactions*. NeurIPS. https://arxiv.org/abs/1706.08566

Schwaller, P., Laino, T., Gaudin, T., Bolgar, P., Hunter, C. A., & Bekas, C. (2019). Molecular transformer: A model for uncertainty-calibrated chemical reaction prediction. *ACS Central Science*, 5(9), 1572-1583.

Schwaller, P., Laino, T., Gaudin, T., Bolgar, P., Hunter, C. A., & Reymond, J.-L. (2018). Molecular transformer: A model for uncertainty-calibrated chemical reaction prediction. *ACS Central Science*, 5(9), 1572–1583.

Schwaller, P., Petraglia, R., Zumbach, J., Courty, M., Blaudeau, A., & Laino, T. (2020). Unassisted and Explainable Prediction of Organic Reactions. *Chemical Science*, 11(2), 455–463. DOI: 10.1039/C9SC04069G.

Segler, M. H. S., Preuss, M., & Waller, M. P. (2018). Planning Chemical Syntheses with Deep Neural Networks and Symbolic AI. *Nature*, 555(7698), 604–610. DOI: 10.1038/nature25978.

Segler, M. H., & Waller, M. P. (2017). Neural-symbolic machine learning for retrosynthesis and reaction prediction. Chemistry – A European Journal, 23(25), 5966–5971. DOI: 10.1002/chem.201605499.

Sun, Q., et al. (2018). "PySCF: The Python-based simulations of chemistry framework." *WIREs Computational Molecular Science*, 8(1), e1340.

Toyao, T., Maeno, Z., Takakusagi, S., Kamachi, T., Takigawa, I., & Shimizu, K. I. (2020). Machine Learning for Catalysis Informatics: Recent Applications and Prospects. *ACS Catalysis*, 10(4), 2260–2297. DOI: 10.1021/acscatal.9b04688.

Toyao, T., Suzuki, K., Kikuchi, S., Takakusagi, S., Shimizu, K. I., & Takigawa, I. (2018). Machine Learning for Catalysis Informatics: Open Challenges and Perspectives. *Catalysis Today*, 310, 98–110. DOI: 10.1016/j.cattod.2018.05.053.

Trott, O., & Olson, A. J. (2010). AutoDock Vina: Improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading. *Journal of Computational Chemistry*, 31(2), 455–461.

Vamathevan, J., et al. (2019). Applications of machine learning in drug discovery and development. *Nature Reviews Drug Discovery*, 18(6), 463-477.

Vaswani, A., Shazeer, N., Parmar, N., et al. (2017). *Attention is all you need*. NeurIPS. https://arxiv.org/abs/1706.03762

Wu, S., & Zhang, L. (2018). AI-powered predictive modeling in toxicology: Recent advances and new horizons. *Frontiers in Artificial Intelligence*, 1, 2.

Wu, Z., et al. (2018). "MoleculeNet: A benchmark for molecular machine learning." *Chemical Science*, 9(2), 513–530. https://deepchem.io

Wu, Z., Ramsundar, B., Feinberg, E. N., Gomes, J., Geniesse, C., Pappu, A. S., &Leswing, K., et al. (2018). MoleculeNet: A benchmark for molecular machine learning. *Chemical Science*, 9(2), 513–530. Available at: https://deepchem.io.

Xie, T., & Grossman, J. C. (2018). Crystal Graph Convolutional Neural Networks. *Physical Review Letters*, 120(14), 145301.

Xin Yang, Yifei Wang, Ryan Byrne, Gisbert Schneider, Shengyong Yang. "Concepts of Artificial Intelligence for Computer-Assisted Drug Discovery", Chemical Reviews, 2019

Xu, K., Hu, W., Leskovec, J., &Jegelka, S. (2019). *How powerful are graph neural networks?* International Conference on Learning Representations (ICLR). https://arxiv.org/abs/1810.00826

**Correspondence:shyamal.m68@gmail.com*

Yang, Q., Yan, Z., Wu, Y., Yu, L., & Chen, Y. (2020). Retrosynthesis Prediction with Deep Learning Algorithms. *Chemical Reviews*, 120(19), 11216–11259. DOI: 10.1021/acs.chemrev.0c00277.

Young, J., & Williams, C. K. (2017). Applications of Machine Learning in Polymer Synthesis and Property Prediction. *Macromolecules*, 50(9), 3690–3700. DOI: 10.1021/acs.macromol.7b00139.

**Correspondence:shyamal.m68@gmail.com*