Innovations

Unlocking the Future of Pharmaceuticals: AI Techniques in Drug Discovery

Dr. Chandan Adhikari
Assistant Professor
Department of Basic Science and Humanities
Institute of Engineering & Management, Salt Lake, Kolkata
University of Engineering & Management, Kolkata
Email: chandan.adhikari@iem.edu.in

Abstract: *The pharmaceutical industry stands on the precipice of a transformative era, with artificial intelligence (AI) techniques playing a pivotal role in reshaping drug discovery processes. This article delves into the cutting-edge AI techniques that are revolutionizing pharmaceutical research, including Generative Adversarial Networks (GANs), Recurrent Neural Networks (RNNs), Reinforcement Learning (RL), and Variational Autoencoders (VAEs). By harnessing the power of these AI tools, the pharmaceutical sector is poised to accelerate drug development, reduce costs, and bring life-changing therapies to patients faster than ever before. Generative Adversarial Networks (GANs) have emerged as a promising AI technique in drug discovery. GANs facilitate the generation of novel molecules with desired properties, enabling researchers to explore a vast chemical space efficiently. By training a generator and discriminator network to compete with each other, GANs can create molecular structures that hold the potential to become breakthrough drugs. Recurrent Neural Networks (RNNs) have found application in understanding the intricate relationships within biological data. With their ability to process sequential data, RNNs can analyze genetic sequences, predict protein structures, and identify potential drug targets. The ability to decipher complex biological information at scale accelerates the identification of novel drug candidates. Reinforcement Learning (RL) brings reinforcement learning techniques into the world of drug discovery. RL algorithms can optimize drug design by learning from trial-and-error simulations. Researchers can use RL to fine-tune drug properties, such as binding affinity or bioavailability, leading to more effective and safer medications. Variational Autoencoders (VAEs) have gained traction for their ability to generate novel molecular structures while preserving important chemical features. By encoding and decoding molecular representations, VAEs can help researchers explore chemical space more systematically. This not only accelerates drug discovery but also enables the design of tailored medications for specific patient populations. AI techniques also play a critical role in drug repurposing, a cost-effective strategy to identify new therapeutic uses for existing drugs. By analyzing vast datasets, AI algorithms can uncover hidden connections between drugs and diseases, potentially fast-tracking the development of treatments for previously unaddressed conditions.*

Introduction

In the quest to combat an ever-growing array of diseases and medical conditions, the pharmaceutical industry has been at the forefront of research and innovation. Traditional drug discovery methods have historically relied on time-consuming and resource-intensive processes. The development of new drugs through these conventional approaches can take many years and often demands substantial financial investment. However, in recent years, advancements in artificial intelligence (AI) have shown promising potential to revolutionize

the drug discovery process. Generative AI, in particular, has emerged as a powerful tool, enabling the generation of novel drug molecules with significant applications in the pharmaceutical industry. The traditional drug discovery process is a long and arduous journey, characterized by extensive experimental procedures, data analysis, and clinical trials.[1, 2] On average, it takes around 10 to 15 years for a drug to transition from the early stages of discovery to final approval for patient use. The cost associated with this process is staggering, often reaching billions of dollars for a single drug candidate. A significant portion of these expenses is attributed to the high attrition rates, where most potential drug molecules fail to progress beyond the preclinical and clinical testing phases due to toxicity, lack of efficacy, or unforeseen adverse effects. Moreover, the reliance on trial and error-based experimentation in conventional drug discovery poses substantial challenges. The vast chemical space of potential drug candidates demands exhaustive screening, leading to an exorbitant number of compounds that must be synthesized and tested. As a result, the process becomes increasingly time-consuming and cost-prohibitive, impeding the timely development of much-needed therapeutics.[3, 4]

Traditional drug discovery faces numerous hurdles that hinder the efficient identification and development of therapeutic agents. One of the primary challenges lies in the complex nature of diseases. Many conditions, especially chronic and multifactorial disorders, have intricate underlying mechanisms that are not fully understood. This lack of comprehensive knowledge often results in the development of drugs that provide only partial relief or produce undesirable side effects. Furthermore, the slow pace of traditional drug discovery limits the rapid response to emerging infectious diseases or rapidly mutating pathogens. This issue becomes evident during global health crises, such as the COVID-19 pandemic, where expediency in drug development is of paramount importance. Another major challenge is the need to consider the vast chemical space when exploring potential drug candidates. Identifying molecules that exhibit the desired therapeutic properties while simultaneously adhering to pharmacological and safety criteria is a daunting task. Traditional methods can only explore a fraction of this vast chemical space, leaving a treasure trove of potential drug molecules undiscovered. Given the challenges and limitations of conventional drug discovery, there is an urgent need to explore alternative methods that can accelerate the identification of potential therapeutic agents.[5, 6] With the growing availability of vast and diverse datasets, there is an opportunity to leverage AI technologies to navigate the complex drug discovery landscape more efficiently. Generative AI, a subset of AI, is particularly promising in addressing these challenges. Generative models, such as Generative Adversarial Networks (GANs) and Variational Autoencoders (VAEs), can learn from existing chemical structures and propose new molecules that exhibit desirable drug-like properties. This approach holds the potential to streamline the drug discovery process by suggesting novel compounds that could be candidates for further evaluation.By integrating generative AI models with advanced computational simulations and virtual screening techniques, researchers can rapidly assess thousands of generated molecules, significantly increasing the scope of potential drug candidates. These alternative methods have the potential to bridge the gap between drug discovery and development, expediting the timeline for delivering novel therapeutics to patients in need.[7]

The advent of AI technologies has brought about a paradigm shift in various industries, including healthcare and pharmaceuticals. In recent years, AI has emerged as a powerful ally in drug discovery, transforming the way researchers approach the identification and development of potential therapeutic agents. AI's integration in drug discovery encompasses multiple aspects, such as data mining, predictive modeling, and image analysis. Machine learning algorithms can efficiently analyze vast datasets, including biological information, chemical structures, and clinical trial data. By detecting patterns and relationships within these datasets, AI models can identify potential drug targets, predict drug efficacy, and optimize compound properties. Generative AI, a subset of machine learning, has shown exceptional promise in generating novel drug molecules. By leveraging generative models, researchers can create virtual chemical libraries with vast diversity, effectively expanding the search space for potential drug candidates. Moreover, generative AI can design molecules with specific properties, such as increased potency, reduced toxicity, or improved

pharmacokinetic profiles. However, the adoption of AI in drug discovery is not without its challenges. The integration of AI-generated drug molecules into the regulatory framework, ensuring their safety and efficacy, remains a critical concern. Robust validation processes and rigorous testing are imperative to establish the reliability of AI-generated molecules for human use.[8] Despite these challenges, the transformative potential of AI in drug discovery is undeniable. The amalgamation of human expertise with the computational power of AI has the capacity to accelerate drug discovery, leading to the development of more effective and personalized therapies, ultimately benefiting patients and advancing healthcare.[9]The introduction of AI-generated drug molecules marks a pivotal moment in the pharmaceutical industry. The conventional drug discovery process, although effective, is time-consuming, costly, and fraught with challenges. The integration of generative AI models presents an opportunity to overcome these hurdles by significantly expediting the identification of potential therapeutic agents. Through the collaborative efforts of researchers, clinicians, and AI technologies, the development of novel and more effective drugs is on the horizon, offering hope for improved patient outcomes and a brighter future in healthcare. As the field continues to evolve, it is crucial to address the ethical, regulatory, and safety considerations surrounding AI-generated drug molecules, ensuring responsible and effective application in the pharmaceutical industry.[10, 11]

The application of AI techniques in drug discovery has introduced new avenues for accelerating the identification of potential therapeutic agents. One such approach is the use of Generative Adversarial Networks (GANs), which consist of two neural networks—the generator and the discriminator—competing against each other. The generator creates novel drug-like molecules, while the discriminator evaluates their authenticity. Through continuous competition and learning, GANs can generate diverse and unique chemical structures, presenting an efficient means of exploring the vast chemical space for potential drug candidates. Recurrent Neural Networks (RNNs) are another powerful AI technique employed in drug discovery. RNNs are capable of capturing sequential data and have been effectively utilized to predict molecular properties, optimize compound designs, and analyze chemical reactions. Their ability to process sequential data makes them well-suited for generating molecular structures with desired properties. Reinforcement Learning (RL) is an AI approach where an agent learns to perform actions in an environment to achieve specific goals. In drug discovery, RL can be employed to optimize drug candidates' properties by iteratively generating and evaluating molecules based on predefined reward functions. This iterative process allows RL to identify molecules with desired properties, enabling the development of novel drug candidates with higher efficacy and reduced side effects. Autoencoders are a type of neural network used for unsupervised learning, primarily focused on data compression and reconstruction. In drug discovery, autoencoders can encode molecular representations and then decode them to generate novel chemical structures. By learning the underlying patterns from existing molecules, autoencoders can propose new molecules that possess similar properties, potentially leading to the discovery of novel drug candidates.[12, 13]

The possibility of discovering drug molecules through AI-generated approaches has ignited new hope in tackling life-threatening diseases. Diseases like cancer, neurodegenerative disorders, and infectious diseases continue to pose significant threats to human health, necessitating novel and effective therapeutics. AI-generated drug molecules offer the advantage of exploring a broader chemical space and identifying molecules with specific properties that could be crucial for combating these diseases. GANs, for instance, have been utilized to generate novel molecules with potent anticancer activity, demonstrating their potential in discovering drug candidates for the treatment of cancer. Furthermore, AI techniques can facilitate the discovery of repurposable drugs—existing drugs that may have therapeutic effects for conditions beyond their original indications. By analyzing large datasets and molecular structures, AI can identify existing drugs with potential applications in the treatment of different diseases, expediting the drug repurposing process. In the context of infectious diseases, AI can play a vital role in rapidly identifying drug candidates with antiviral properties, essential during pandemics or outbreaks. AI-powered models can efficiently screen vast libraries of compounds, leading to the discovery of potential antiviral agents in a shorter time frame, ultimately saving lives.[14, 15]

Over the past five years, the development of AI in drug discovery has witnessed significant strides, with research spanning across various domains. One notable area of progress is in the generation of novel drug-like molecules using GANs and other generative models. AI-powered approaches have demonstrated the potential to accelerate the process of drug discovery by proposing molecules with desired properties, significantly reducing the time and resources required for traditional methods. AI has also proven valuable in predicting the efficacy and safety of drug candidates, optimizing lead compounds, and identifying potential drug targets. By analyzing vast amounts of data, including biological, chemical, and clinical information, AI models can provide insights into the interactions between molecules and biological systems, aiding in the rational design of drugs. Moreover, the integration of AI with high-throughput screening techniques has enabled the screening of vast chemical libraries, expediting the identification of drug candidates with the potential to treat various diseases. This approach has been especially promising in the context of neglected and rare diseases, where the identification of therapeutics is often challenging.[16, 17]

In the rapidly evolving field of AI-generated drug molecules, a comprehensive review article plays a crucial role in consolidating and summarizing the current state-of-the-art research and applications. As AI techniques continue to make significant strides in drug discovery, there is a need to provide a structured and accessible overview of the latest developments, methodologies, and challenges. A review article can serve as a valuable resource for researchers, pharmaceutical professionals, and policymakers, offering insights into the potential impact of AI-generated drug molecules on the pharmaceutical industry. Additionally, it can highlight the unique opportunities and limitations of each AI technique, aiding researchers in selecting appropriate methods for their specific drug discovery needs. Furthermore, by critically assessing the advancements and discussing the ethical considerations associated with AI-generated drug molecules, a review article can guide responsible and transparent AI integration in drug development. This, in turn, can foster collaboration and knowledge-sharing among researchers and promote the responsible use of AI in the pharmaceutical industry. Overall, a well-structured and informative review article holds immense value in advancing the field of AI in drug discovery, propelling the development of novel therapeutics, and ultimately contributing to the improvement of global healthcare.[18-20]

Various AI Techniques for Drug Discovery and Molecule Generation

Generative Adversarial Networks (GANs) are aiding drug discovery by generating novel molecular structures with desired properties, enabling the rapid exploration of chemical space and the synthesis of potential drug candidates. Recurrent Neural Networks (RNNs) play a crucial role in predicting molecular properties, bioactivity, and toxicity, providing valuable insights into potential drug efficacy and safety profiles. Reinforcement Learning (RL) optimizes drug candidate selection by guiding molecular design through trial-and-error simulations, effectively accelerating the process of identifying promising compounds. Variational Autoencoders contribute to drug discovery by transforming molecular representations into a continuous latent space, allowing for efficient exploration of chemical structures and assisting in the identification of lead compounds with desired characteristics. The integration of these techniques facilitates a more efficient and data-driven drug discovery process, potentially revolutionizing the pharmaceutical industry with faster and more effective drug development.[21, 22] In below a brief overview of each techniques for drug discovery and molecule generation are discussed from the last five years development.

Generative Adversarial Networks (GANs):

Huang et al. introduces a generative model called DeepGAN, based on the Generative Adversarial Network algorithm. The model utilizes DeepSMILES as a training object, overcoming limitations associated with SMILES. Additionally, reinforcement learning is incorporated to handle non-differentiable problems in the discriminator. The model is trained to optimize rewards and adversarial loss, resulting in superior performance compared to other tested models like ORGAN, OR(W)GAN, and Naive RL. Experimental results demonstrate that DeepGAN generates diverse and valid molecules while improving desired metrics in the

drug discovery process. [23] Singhal et al. highlights the limitations of deep learning methods due to the requirement for a large number of high-quality data samples. To address this, the authors propose a generative modeling based computational framework using three variants of Generative Adversarial Network (GAN) to synthesize images for phenotypic profiling of drug-induced perturbations. They find that the Deep Convolutional GAN (DCGAN) is the most efficient in generating realistic synthetic images. A pre-trained convolutional neural network (CNN) is used to extract features from both real and synthetic images, and a classification model is trained on both datasets. The quality of synthesized images is evaluated by comparing their feature distributions with real images, and the DCGAN-generated images are used to augment the real image dataset, resulting in improved classification performance. The proposed method is also demonstrated on the generation of bacterial images and their feature distributions for different drug treatments. Overall, the results show that the DCGAN-based framework enables the generation of realistic synthetic high-content images, facilitating the study of drug-induced effects on cells and bacteria.[24] The study by Kotsias et al. focuses on utilizing deep learning methods for drug discovery through the creation of novel molecular structures. A new deep learning architecture called "LatentGAN" is proposed, which combines an autoencoder and a generative adversarial neural network for de novo molecular design. The method is applied in two scenarios: generating random drug-like compounds and generating target-biased compounds. The results demonstrate the effectiveness of the LatentGAN method in both cases. Compounds sampled from the trained model occupy a similar chemical space as the training set and yield a substantial number of novel compounds. Additionally, the drug-likeness score of generated compounds matches that of the training set. Comparing LatentGAN to a Recurrent Neural Network-based generative model approach reveals that both methods can be used complementarily, as they produce different sets of compounds. Overall, this research demonstrates the potential of deep learning for advancing drug discovery efforts.[25] A major challenge in AI-based drug design is identifying which molecules should be prioritized for synthesis and biological evaluation, as the trial-and-error process remains resource-intensive. Tong et al. introduces a novel molecular filtering method called MolFilterGAN, based on a progressively augmented generative adversarial network, to address this challenge. The research demonstrates that traditional screening metrics fail to differentiate AI-designed molecules effectively. In contrast, MolFilterGAN outperforms conventional screening approaches based on drug-likeness or synthetic ability metrics. Retrospective analysis of AI-designed inhibitors for the discoidin domain receptor 1 (DDR1) shows that MolFilterGAN significantly improves the efficiency of molecular triaging. The evaluation of MolFilterGAN on eight external ligand sets further demonstrates its effectiveness in enriching bioactive compounds across various target types. The results emphasize the importance of MolFilterGAN in evaluating molecules comprehensively and accelerating molecular discovery, especially when combined with advanced AI generative models.[26] A deep convolutional generative adversarial network (dcGAN) model was developed by Xiang et al. to design novel compounds targeted for cannabinoid receptors. The model consists of two components, the discriminator D and the generator G, which are trained in an adversarial process. D learns to distinguish between authentic compounds and "fake" compounds generated by G, while G aims to optimize its weights to produce "fake" compounds that can fool D. To determine the best architecture and input data structure for the convolutional neural networks (CNNs) involved, various combinations of network architectures and molecular fingerprints were explored. CNN models like LeNet-5, AlexNet, ZFNet, and VGGNet were investigated, and four types of fingerprints (MACCS, ECFP6, AtomPair, and AtomPair Count) were calculated to represent the diverse structural characteristics of small molecules. While generating fingerprints as output has limitations in directly converting them into concrete molecular structures, the generative models with convolutional networks show promising opportunities for molecule screening and rational modifications in computer-aided drug discovery. The study highlights how recent advances in deep learning can benefit computer-aided drug discovery. [27] Due to the sensitive and geo-distributed nature of molecular data, pharmaceutical companies are often unwilling or unable to share their local datasets for centralized training of GANs. To address this issue, Xiang et. al. proposes a novel framework called GraphGANFed (Graph convolutional network in Generative Adversarial Networks via Federated

learning). This framework combines graph convolutional neural networks (GCNs), GANs, and federated learning (FL) to generate novel molecules without sharing local datasets. The discriminator is implemented as a GCN to better capture molecular features represented as graphs, and FL is used to train both the discriminator and generator in a distributed manner, preserving data privacy. Extensive simulations based on three benchmark datasets are conducted to demonstrate the feasibility and effectiveness of GraphGANFed. The results show that molecules generated by GraphGANFed exhibit high novelty and diversity. The simulations also reveal insights such as the importance of using a lower complexity discriminator for smaller datasets to avoid mode collapse, the tradeoff among different evaluation metrics, and the significance of employing the right dropout ratio for the generator and discriminator to prevent mode collapse.[28] The study by Surana et. al. addresses the urgent need for efficient drug development due to the continuous rise in pathogenic viruses. Antiviral peptides (AVPs) are gaining attention as potential drug candidates, but their diverse sequences and limited characterization hinder their applications. To expedite the process of developing novel peptide drugs, the researchers developed PandoraGAN, an advanced deep learning approach that uses a curated dataset of 130 highly active peptides. PandoraGAN generates novel antiviral peptides by learning implicit properties from the training data and validates them based on physico-chemical properties. The generated sequences are compared with the training dataset, confirming PandoraGAN's capability to produce novel antiviral peptide backbones similar to known active peptides. This method presents a potential for discovering previously unseen AVP patterns, and it represents the first use of GAN models for antiviral peptides across the viral spectrum.[29] Hou et al. introduces a novel algorithm called ChemistGA, which combines traditional heuristic algorithms with Deep Learning (DL) techniques. ChemistGA redefines the crossover process of the traditional genetic algorithm (GA) using DL in conjunction with GA, and incorporates an innovative backcrossing operation to create desired molecules. The results demonstrate that ChemistGA not only preserves the strengths of the traditional GA but also significantly improves the ability to generate molecules with desired properties. Comparative evaluations against state-of-the-art baselines on two benchmarks highlight ChemistGA's impressive performance and its potential to revolutionize real-world drug discovery scenarios through the application of generative models.[30]

Recurrent Neural Networks:

Recurrent Neural Networks (RNNs) play a crucial role in drug delivery by predicting and optimizing drug properties and interactions. RNNs can analyze complex sequential data, such as molecular structures and pharmacological activities, to design better drug delivery systems. By learning patterns and relationships in the data, RNNs aid in predicting drug solubility, bioavailability, toxicity, and target interactions. This enables researchers to identify potential drug candidates more efficiently, leading to improved drug delivery methods and enhanced therapeutic outcomes.[31]

Waller et al.introduces the application of computational strategies in de novo drug design, where novel molecules are generated with high affinity to specific biological targets. The study explores the use of recurrent neural networks as generative models for molecular structures, akin to statistical language models in natural language processing. The research demonstrates that the generated molecules exhibit properties that closely correlate with those used to train the model. To enhance libraries with molecules active against specific biological targets, the authors propose fine-tuning the model using small sets of known active molecules. The model was tested against Staphylococcus aureus and Plasmodium falciparum (Malaria), successfully reproducing a significant percentage of test molecules designed by medicinal chemists. When combined with a scoring function, the model enables a complete de novo drug design cycle, facilitating the generation of extensive sets of novel molecules for drug discovery.[14] The work done by Yasonik et al. proposes a novel de novo approach for optimizing multiple traits of molecules collectively. It uses a recurrent neural network to generate molecules, which are then ranked based on multiple properties using a nondominated sorting algorithm. The best molecules are selected and used to fine-tune the neural network through transfer learning, creating a cycle that mimics the traditional design-synthesis-test cycle. The

approach is demonstrated through a proof of concept, where it optimizes multiple molecular properties simultaneously. After five iterations of the cycle, there was a 14-fold improvement in the quality of generated molecules, improved accuracy of the neural network, and increased structural diversity of the molecules, all without requiring large amounts of training data or handwritten scoring functions. This approach uniquely combines scalable generation with multiobjective optimization of molecules. [32] Yao, H. et al. explores the use of artificial intelligence (AI) in drug discovery, specifically focusing on de novo molecular generation using recurrent neural networks. The aim is to discover new chemical space for kinase inhibitors, an important area in medicinal chemistry. The researchers successfully generated one potent Pim1 inhibitor and two lead compounds that inhibit CDK4, demonstrating the potential of AI-based molecular generation in drug development. The study highlights the importance of novelty in highly competitive medicinal chemistry fields and suggests that AI can provide valuable insights and practical applications in drug discovery.[33] Santhosh et al. propose a new drug discovery method that utilizes LSTM models to generate novel molecules capable of binding to the novel Coronavirus protease. The study shows that the method can successfully create molecules similar to trained ones. They fine-tune the model to generate drug-like molecules targeting the 3CLPro protease, an important therapeutic target for Covid-19. In silico screening reveals that 80% of the generated molecules have strong binding affinities, with the top candidate showing a significantly better binding score than approved commercial drugs like Remdesivir. This suggests that the generated molecules have potential as Covid-19 drug candidates.[34] Bjerrum et al. propose a simple approach for focused molecular generation in drug design using a conditional recurrent neural network (cRNN). They used selected molecular descriptors to initialize the network's memory state, and then the cRNN generates alphanumeric strings describing molecules. This allows them to address the inverse design problem directly, generating molecules meeting specified conditions. They also introduce a method to assess the focus of the model's conditional output using negative log-likelihood plots. The cRNN's output is more focused than traditional unbiased RNNs but less focused than autoencoders, providing an intermediate output specificity. The proposed architecture shows promise for steering sequential data generation with recurrent neural networks.[35] In a work conducted by Santos et al. shows an evaluation on Recurrent Neural Networks which can learn the syntax of molecular representation in terms of SMILES notation. We optimize the computational framework based on the recurrent architecture and its hyper-parameters. Moreover, we evaluate the performance of two types of encoding and spatial arrangement of molecules: Embedding and One-hot Encoding, and datasets with and without stereo-chemical information, respectively. The proposed model showed improved performance when compared to the current literature, both in terms of percentage of valid generated SMILES and diversity with 98.7% and 0.88, for the ChEMBL dataset, respectively. Even when considering the ZINC biogenic library, with stereochemical information, the values were 94.5% and 0.90. The obtained results reveal the potential of the recurrent architectures in learning the SMILES syntax and adding novelty to generate promising compounds.[36] The study done by Srinivasa et al. explores the use of memory-augmented RNN-based architectures (Neural Turing Machine and Differentiable Neural Computer) for generating small molecules. They use a character-level CNN to predict molecule properties and employ deep reinforcement learning to guide molecule generation towards desired properties. The research compares the performance of these architectures with simpler RNNs (Vanilla RNN, LSTM, and GRU) to understand the impact of memory augmentation on de-novo drug generation in terms of validity, novelty, and property bias.[37] Yang et al. proposed MGRNN (Molecular Graph Recurrent Neural Networks) which is a graph recurrent neural network model for generating drug molecular structures. It combines the benefits of iterative molecular generation algorithms and efficient training strategies. MGRNN demonstrates efficient computation during training, high model robustness for data, and an iterative sampling process that allows for valency checking using chemical domain expertise. Experimental results indicate that MGRNN can generate 69% chemically valid molecules even without chemical knowledge and 100% valid molecules when incorporating chemical rules.[38] Falcao et al. investigated the impact of applying quantization during training on Recurrent Neural Networks (RNNs) used for SMILES generation in drug discovery. The study

compares three commonly used RNN algorithms (Simple RNN, LSTM, and GRU) and tests various quantization configurations using the QK-eras library. The goal is to generate a large number of novel SMILES to accelerate the drug discovery process. The results indicate that LSTM and GRU layers perform well with 4-bit quantization, while quantizing Simple RNN does not yield significant improvements. Understanding the behavior of quantized networks helps control the efficiency of the model selection and quantization process.[39] Anguang et al. proposed a framework which uses Recurrent Neural Networks (RNN) with an attention model to sample the chemical space of organophosphorus molecules through a fragment-based approach. It is trained on a ZINC dataset with high drug likeness scores. The objective is to predict molecules with similar biological action to organophosphorus pesticides or chemical warfare agents but with reduced human toxicity. The generated molecules have a starting fragment of PO2F and a bulky hydrocarbon side chain that limits their binding effectiveness to the targeted protein.[40] Hoffmann et al. proposed a research project aiming to create new chemical structures with an affinity to specific protein domains using a deep neural network. They utilized SELFIES codes to transfer chemical information to the neural network, allowing for the generation of novel compounds. The generated structures were filtered based on drug-likeness criteria and synthetic accessibility. The affinity to selected protein domains was verified using the AutoDock tool. The study successfully identified chemical structures with an affinity to protein domains with PDB IDs 7NPC, 7NP5, and 7KXD.[41]

Reinforcement Learning (RL)

Reinforcement Learning (RL) is being applied to drug molecule generation to optimize the discovery of new and effective pharmaceutical compounds. RL algorithms guide the process by learning from iterative interactions with molecular simulations and scoring functions. They generate and modify molecular structures, aiming to maximize desired properties such as binding affinity, bioavailability, and safety, while minimizing undesirable characteristics. This approach accelerates drug discovery by exploring vast chemical space more efficiently, potentially leading to faster identification of promising drug candidates.[42] Popova et al. have introduced ReLeaSE (Reinforcement Learning for Structural Evolution), a novel computational strategy for designing molecules with specific properties. It combines generative and predictive deep neural networks trained separately and then jointly using reinforcement learning. The method employs simplified molecular-input line-entry system (SMILES) representations for molecules. Generative models create feasible SMILES strings, while predictive models forecast desired properties. The process involves initial separate supervised training, followed by joint training using reinforcement learning to bias new molecule generation towards desired properties. The approach was successfully applied to design chemical libraries targeting various properties and biological activities. This method has broad potential for generating tailored chemical libraries with optimized properties.[43] Ribeiro, B. et al. explores using deep learning to generate potential new drugs by creating molecules with specific biological properties. It uses two neural networks: a Generator that makes valid molecules using SMILES notation, and a Predictor that evaluates the molecules' affinity for a target. The Generator is refined through Reinforcement Learning, using an innovative strategy that involves two Generators to enhance novelty in generated compounds. This strategy balances exploring new chemical space and using existing knowledge. The method is tested by designing molecules with specific properties, showing successful results in generating diverse and promising molecules.[44]Mercado et al. introduces a new method that employs reinforcement learning to enhance graph-based deep generative models for creating new molecules. The approach successfully guides a pre-trained model to generate molecules with desired properties, even if those properties were not present in the training data. The study tackled tasks like generating smaller/larger molecules, improving drug-likeness, and increasing bioactivity. The results demonstrated that the proposed method outperforms previous techniques in generating diverse compounds with predicted DRD2 activity, achieving a 95% success rate.[45] Shapira, B. et al. introduces Taiga, a transformer-based model for generating molecules with specific properties. It utilizes a two-stage approach: first predicting molecules' next tokens using language modeling, then using reinforcement learning to

optimize properties like QED. Taiga shows strong performance, outperforming existing methods in molecule optimization by up to 20%, demonstrated across different datasets and tasks. It generates molecules with improved biological property scores compared to the same model without reinforcement learning.[46] Yang et al. introduces DRlinker, a novel framework for fragment-based drug discovery. DRlinker employs reinforcement learning to guide the linking of fragments, creating compounds with specified attributes. The approach is effective in tasks such as controlling linker length and log P, optimizing predicted bioactivity, and achieving multiple objectives. The model successfully generated compounds meeting desired linker length and log P in high percentages, improved bioactivity optimization, and facilitated scaffold hopping while maintaining 3D dissimilarity from lead inhibitors. DRlinker shows promise in practical fragment-based drug design.[47] Ishitani et al. proposed a new approach for automatically designing molecules with specific chemical and biochemical properties. The researchers developed a reversible tree representation of molecules called "Reversible Junction Tree" (RJT), which can be converted back into the original molecule. They used deep reinforcement learning (RJT-RL) to construct molecules as a tree structure, ensuring that all intermediate and final states are valid molecules. This method efficiently guides molecule optimization in simple tasks and is also applicable to more complex tasks like multiobjective optimization and fine-tuning in drug discovery.[48] Pereira et al. explores the use of Deep Learning techniques in early drug discovery to address the challenges of time and cost. The researchers trained a recurrent neural network to generate molecules with desired properties using SMILES strings and optimized it through Reinforcement Learning. A second neural network assessed the fitness of generated molecules. The model successfully designed molecules with enhanced affinity for a specific receptor, maintaining validity and diversity. This approach holds promise for accelerating drug development.[49] Kumari, D. et al. introduced graph-based deep learning for designing potential therapeutic drugs against SARS-CoV-2. The method involves two components: a novel reinforcement learning-based graph generator with a knowledge graph, and a fusion approach for predicting binding strength. The generator employs a gated graph neural network and knowledge graph for compound creation, while the fusion approach estimates binding affinity between generated molecules and proteins. Experiments demonstrate successful refinement of generated molecules and efficient molecule screening, resulting in promising compounds against SARS-CoV-2 protein. The study achieved notable binding affinity and compared generated compounds with an existing drug, Indinavir, for drug development insights.[50]Priyakumar et al. proposes a novel approach using reinforcement learning to generate molecules with strong binding to the target and favorable drug-like properties. A deep generative model is trained to create drug-like molecules, then optimized with reinforcement learning to produce molecules with desired characteristics such as LogP, drug likeliness, surface area, and binding affinity. The study introduces a unique strategy to periodically change the reward calculation for multi-objective optimization, outperforming conventional methods and generating more molecules with desired properties.[51]Li, L. et al. studied the MolDQN framework which combines chemistry expertise and advanced reinforcement learning techniques (double Q-learning and randomized value functions) for molecule optimization. It directly modifies molecules while ensuring chemical validity, without pre-training on any dataset to avoid bias. MolDQN performs as well as or better than recent algorithms on benchmark tasks, though these tasks may not represent real drug discovery challenges. To address this, MolDQN incorporates multi-objective reinforcement learning to optimize drug-likeness and molecular similarity. The framework is demonstrated by optimizing molecules and revealing the optimization path in chemical space.[52]

Variational Autoencoders

Variational Autoencoders (VAEs) play a crucial role in drug delivery and molecule generation by enabling the efficient design and discovery of new molecules with desired properties. VAEs leverage their ability to learn complex patterns from molecular data to encode the molecular structure and properties into a latent space. This latent space representation allows for systematic exploration and manipulation of chemical space, facilitating the identification of molecules that exhibit optimal characteristics for drug delivery, such as

solubility, bioavailability, and target specificity. By generating novel molecular structures and predicting their properties, VAEs expedite the process of molecular design, lead optimization, and the discovery of innovative therapeutic agents, thereby accelerating advancements in drug development and personalized medicine.[53] Min et al. has created molecules with specific properties for applications like drug development and organic materials. It proposes a model, MGCVAE, based on an autoencoder, to generate molecules. The performance of MGCVAE was compared to another model, MGVAE. MGCVAE successfully generated molecules with desired properties, showing a significant improvement over MGVAE. The study also used multi-objective optimization to design molecules with two properties simultaneously. The results indicate that MGCVAE is effective in generating molecules that meet specific physical property criteria, suggesting the potential of data-driven models for designing new molecules.[54]The study by Waldispühl et al. introduces a novel approach for drug design using deep generative models. Unlike existing methods, this approach combines information from the structure of the target molecule and the chemical space. By using a ligand-centered generative model, the researchers iteratively create molecules that fit the target structure better, guided by molecular docking simulations. A new graph-to-Selfies Variational Autoencoder (VAE) is proposed, which significantly speeds up the process of decoding while maintaining performance. Through this approach, they successfully improve the docking scores of generated molecules, leading to a substantial enrichment of high-scoring compounds compared to traditional methods.[55]Zhu et al. studied Sc2Mol, a generative model named for molecule synthesis. It operates using SMILES strings for molecules and comprises two steps: scaffold generation through a variational autoencoder and scaffold decoration via a transformer. This approach proves effective for both random molecule creation and scaffold enhancement. The model demonstrates success in learning distribution patterns and molecule optimization within drug-like datasets. Additionally, Sc2Mol autonomously learns rules for refining scaffolds into advanced drug candidates, aligning with established lead optimization principles.[56]The paper published by Zhao et al. showed MoVAE, a new approach for generating molecules to expedite drug discovery. MoVAE employs variational autoencoders (VAEs) to learn and represent molecular structures effectively. Unlike previous VAE-based methods that involve complex graph matching and tend to produce invalid molecules, MoVAE encodes and decodes individual nodes and edges without matching. It enhances molecule validity through adversarial training of the encoder and decoder, acting as generator and discriminator. MoVAE also includes drug property and valence histogram constraints to create molecules meeting specific conditions. Experimental results on real datasets demonstrate MoVAE's superiority over existing algorithms in terms of performance.[57]The study by Kumar et al. involves molecule generation, which is the creation of new chemicals with specific properties. This is achieved by encoding chemicals as continuous vectors and using a variational autoencoder with gated recurrent unit cells for decoding. These cells control the model's complexity. The resulting variational autoencoder achieves a 92.32% validity rate and an 89.63% reconstruction accuracy, outperforming other techniques. The model is effective for generating varied chemical compounds.[58]Nakamura et al. compares the effectiveness of two methods, chemical variational autoencoder (VAE) and similarity search, for generating new functional molecules. Using natural porphyra-334 as a model, three groups of molecules were generated: using mycosporine-like amino acids (MAAs) as seeds (GSEEDS), chemical VAE (GVAE), and similarity search (GSIM). GSEEDS produced 52 molecules meeting porphyra-334's light absorption criteria, GVAE produced 138, and GSIM produced 6. Chemical VAE, utilizing quantum chemistry wave function properties, led to promising molecular designs comparable to porphyra-334, some with unexpected geometries. The study concludes by showcasing a group of molecules discovered through this method.[59]Zhang et al. introduces GF-VAE, a novel approach for generating molecules with desired properties in drug discovery. It combines a flow-based variational autoencoder (VAE) with a lightweight flow model as its decoder. This design speeds up training by optimizing both the encoder and decoder simultaneously. The model leverages the invertibility of the flow model for efficient molecule generation and ensures validity through correction. GF-VAE performs well on various tasks, including molecule generation, reconstruction, latent space smoothness, and property optimization. It outperforms existing methods,

achieving significant improvements in terms of time performance on classical datasets.[60]The study by Verkhivker et al. introduces a new approach to autonomously design molecules that inhibit protein kinases. This method combines machine learning techniques, like variational autoencoders, to map molecules into a latent space and a cluster-based perturbation approach to navigate that space efficiently. This enables the generation of diverse molecules with desired properties. A classifier guides the optimization of these molecules for kinase inhibition. The approach clusters similar molecules together, allowing smooth changes between them. The results show that the strategy effectively explores the molecular space, producing novel kinase inhibitors with high potential for inhibiting specific kinases. This suggests that tailoring latent spaces for specific tasks can enhance autonomous chemical design.[61]Ahn et al. studied a reinforcement learning model that enhances the binding affinity between generated molecules and target proteins. The model employs a Stacked Conditional Variation AutoEncoder (Stack-CVAE) to create molecules with desired chemical traits and strong binding to specific proteins. By using sorafenib's properties and target kinases, the study produced 1000 unique chemical formulas. Stack-CVAE outperformed other generative models, yielding valid compounds with superior binding affinity. In-depth analysis of the top 100 compounds highlighted their novelty, especially for Raf kinases, along with high druggability and synthesizability.[62]The study by Piyayotai et al. explores the impact of chemical compounds and therapies on gene transcription, crucial for clinical and research applications. A generative model, BiCEV, is introduced, using an autoencoder approach to create new molecules based on gene expression. BiCEV effectively generates molecules with high validity (96%), uniqueness (98%), and diversity (0.77). The model's potential is tested on gene-knockdown profiles and drug pair combinations, showing promising results in molecular design quality. While functional equivalence assessment yielded mixed outcomes, the model shows promise for aiding drug discovery by supporting early hit identification and lead optimization with further development and in vitro validation.[63]

Conclusion and Future Aspects

The integration of artificial intelligence (AI) techniques into drug discovery has marked a transformative shift in the pharmaceutical industry. This convergence of cutting-edge technology and biomedical research has the potential to accelerate the development of novel therapeutics, streamline the drug discovery process, and enhance the overall efficacy and safety of pharmaceutical products. In this article, we explored several AI techniques, including Generative Adversarial Networks (GANs), Recurrent Neural Networks (RNNs), Reinforcement Learning (RL), and Variational Autoencoders (VAEs), that are at the forefront of this exciting revolution. One of the key takeaways from this exploration is the versatility of AI techniques in drug discovery. GANs have shown promise in generating molecular structures and designing novel compounds, offering a creative approach to discovering potential drug candidates. RNNs have proven their effectiveness in analyzing sequential data, such as genetic information, aiding in the identification of disease markers and targets. RL, with its ability to optimize complex processes, can enhance the drug development pipeline by guiding experimental design and optimization. VAEs, on the other hand, excel in extracting meaningful representations from large datasets, aiding in the identification of subtle patterns in biological data that may have previously gone unnoticed. The successful application of these AI techniques in drug discovery has already resulted in several breakthroughs.[64-66] AI-powered drug discovery platforms are becoming increasingly adept at predicting drug-target interactions, optimizing chemical compound libraries, and even predicting potential side effects, ultimately reducing the time and cost traditionally associated with bringing a drug to market. Moreover, these techniques have the potential to address some of the most challenging issues in the pharmaceutical industry, such as the development of treatments for rare diseases, where traditional methods may be less effective due to limited data. However, it's crucial to acknowledge that the journey to fully realizing the potential of AI in pharmaceuticals is not without its challenges. The need for high-quality, diverse, and well-curated data remains a critical barrier. Additionally, the interpretability and explainability of AI-driven predictions and decisions in drug discovery are ongoing concerns, particularly in a field where

regulatory approval is paramount. Overcoming these hurdles will require continued collaboration between AI experts, data scientists, and domain-specific researchers. Looking ahead, the future of pharmaceuticals is undeniably intertwined with AI. As technology continues to advance, we can expect to see even more sophisticated AI techniques emerge, capable of tackling increasingly complex drug discovery tasks. Furthermore, the convergence of AI with other transformative technologies, such as CRISPR gene editing and advanced imaging techniques, holds the potential to revolutionize how diseases are treated at the molecular level. The democratization of AI tools and platforms will also play a crucial role in the future of pharmaceuticals. As AI becomes more accessible, smaller biotech companies and academic research institutions will have the opportunity to leverage these technologies to compete with industry giants, fostering a more dynamic and innovative landscape [67-70] In conclusion, AI techniques in drug discovery are ushering in a new era of pharmaceutical research and development. These techniques are poised to enhance our understanding of diseases, streamline drug development processes, and ultimately bring safer and more effective treatments to patients. While challenges persist, the promise of AI in the pharmaceutical industry is undeniable. As we move forward, continued collaboration, data sharing, and ethical considerations will be essential to unlock the full potential of AI in shaping the future of pharmaceuticals. By harnessing the power of AI, we have the opportunity to accelerate the pace of drug discovery and improve the lives of countless individuals worldwide.

References:

1. *Chamberlain, P.P. and L.G. Hamann, Development of targeted protein degradation therapeutics. Nature Chemical Biology, 2019.* ***15****(10): p. 937-944.*
2. *Ward, R.A., et al., Challenges and Opportunities in Cancer Drug Resistance. Chemical Reviews, 2021.* ***121****(6): p. 3297-3351.*
3. *Soltani, S., S. Hallaj-Nezhadi, and M.R. Rashidi, A comprehensive review of in silico approaches for the prediction and modulation of aldehyde oxidase-mediated drug metabolism: The current features, challenges and future perspectives. Eur J Med Chem, 2021.* ***222****: p. 113559.*
4. *Scannell, J.W. and J. Bosley, When Quality Beats Quantity: Decision Theory, Drug Discovery, and the Reproducibility Crisis. PLOS ONE, 2016.* ***11****(2): p. e0147215.*
5. *Scannell, J.W., et al., Diagnosing the decline in pharmaceutical R&D efficiency. Nature Reviews Drug Discovery, 2012.* ***11****(3): p. 191-200.*
6. *Hopkins, A.L., Network pharmacology: the next paradigm in drug discovery. Nature Chemical Biology, 2008.* ***4****(11): p. 682-690.*
7. *Angermueller, C., et al., Deep learning for computational biology. Molecular Systems Biology, 2016.* ***12****(7): p. 878.*
8. *Segler, M.H.S., M. Preuss, and M.P. Waller, Planning chemical syntheses with deep neural networks and symbolic AI. Nature, 2018.* ***555****(7698): p. 604-610.*
9. *Ching, T., et al., Opportunities and obstacles for deep learning in biology and medicine. Journal of The Royal Society Interface, 2018.* ***15****(141): p. 20170387.*
10. *Schwaller, P., et al., Molecular transformer: a model for uncertainty-calibrated chemical reaction prediction. ACS central science, 2019.* ***5****(9): p. 1572-1583.*
11. *Brown, N., et al., GuacaMol: benchmarking models for de novo molecular design. Journal of chemical information and modeling, 2019.* ***59****(3): p. 1096-1108.*
12. *Gómez-Bombarelli, R., et al., Automatic chemical design using a data-driven continuous representation of molecules. ACS central science, 2018.* ***4****(2): p. 268-276.*
13. *Pushpakom, S., et al., Drug repurposing: progress, challenges and recommendations. Nature reviews Drug discovery, 2019.* ***18****(1): p. 41-58.*
14. *Segler, M.H.S., et al., Generating Focused Molecule Libraries for Drug Discovery with Recurrent Neural Networks. ACS Central Science, 2018.* ***4****(1): p. 120-131.*

15. *Aliper, A., et al., Deep learning applications for predicting pharmacological properties of drugs and drug repurposing using transcriptomic data. Molecular pharmaceutics, 2016.* ***13****(7): p. 2524-2530.*
16. *Lavecchia, A., Machine-learning approaches in drug discovery: methods and applications. Drug discovery today, 2015.* ***20****(3): p. 318-331.*
17. *Silver, D., et al., Mastering the game of go without human knowledge. nature, 2017.* ***550****(7676): p. 354-359.*
18. *Stokes, J.M., et al., A deep learning approach to antibiotic discovery. Cell, 2020.* ***180****(4): p. 688-702.*
19. *Olivecrona, M., et al., Molecular de-novo design through deep reinforcement learning. Journal of cheminformatics, 2017.* ***9****(1): p. 1-14.*
20. *Jiménez-Luna, J., F. Grisoni, and G. Schneider, Drug discovery with explainable artificial intelligence. Nature Machine Intelligence, 2020.* ***2****(10): p. 573-584.*
21. *Jiménez-Luna, J., et al., Artificial intelligence in drug discovery: recent advances and future perspectives. Expert Opinion on Drug Discovery, 2021.* ***16****(9): p. 949-959.*
22. *Cerchia, C. and A. Lavecchia, New avenues in artificial-intelligence-assisted drug discovery. Drug Discovery Today, 2023.* ***28****(4): p. 103516.*
23. *Xu, M., et al. Deepgan: Generating molecule for drug discovery based on generative adversarial network. in 2021 IEEE Symposium on Computers and Communications (ISCC). 2021. IEEE.*
24. *Hussain, S., et al., High-content image generation for drug discovery using generative adversarial networks. Neural Networks, 2020.* ***132****: p. 353-363.*
25. *Prykhodko, O., et al., A de novo molecular generation method using latent vector based generative adversarial network. Journal of Cheminformatics, 2019.* ***11****(1): p. 74.*
26. *Liu, X., et al., MolFilterGAN: a progressively augmented generative adversarial network for triaging AI-designed molecules. Journal of Cheminformatics, 2023.* ***15****(1): p. 42.*
27. *Bian, Y., et al., Deep Convolutional Generative Adversarial Network (dcGAN) Models for Screening and Design of Small Molecules Targeting Cannabinoid Receptors. Molecular Pharmaceutics, 2019.* ***16****(11): p. 4451-4460.*
28. *Manu, D., et al., GraphGANFed: A Federated Generative Framework for Graph-Structured Molecules Towards Efficient Drug Discovery. 2023.*
29. *Shraddha, S., et al., PandoraGAN: Generating antiviral peptides using Generative Adversarial Network. bioRxiv, 2022: p. 2021.02.15.431193.*
30. *Wang, J., et al., ChemistGA: A Chemical Synthesizable Accessible Molecular Generation Algorithm for Real-World Drug Discovery. Journal of Medicinal Chemistry, 2022.* ***65****(18): p. 12482-12496.*
31. *Wang, M., et al., Deep learning approaches for de novo drug design: An overview. Current Opinion in Structural Biology, 2022.* ***72****: p. 135-144.*
32. *Yasonik, J., Multiobjective de novo drug design with recurrent neural networks and nondominated sorting. Journal of Cheminformatics, 2020.* ***12****(1): p. 14.*
33. *Li, X., et al., Chemical space exploration based on recurrent neural networks: applications in discovering kinase inhibitors. Journal of Cheminformatics, 2020.* ***12****(1): p. 42.*
34. *Amilpur, S. and R. Bhukya, Predicting novel drug candidates against Covid-19 using generative deep neural networks. Journal of Molecular Graphics and Modelling, 2022.* ***110****: p. 108045.*
35. *Kotsias, P.-C., et al., Direct steering of de novo molecular generation with descriptor conditional recurrent neural networks. Nature Machine Intelligence, 2020.* ***2****(5): p. 254-265.*
36. *Santos, B.P., et al. Optimizing Recurrent Neural Network Architectures for De Novo Drug Design. in 2021 IEEE 34th International Symposium on Computer-Based Medical Systems (CBMS). 2021.*
37. *Suresh, N., et al., Memory augmented recurrent neural networks for de-novo drug design. PLOS ONE, 2022.* ***17****(6): p. e0269461.*
38. *Lai, X., et al., MGRNN: Structure Generation of Molecules Based on Graph Recurrent Neural Networks. Molecular Informatics, 2021.* ***40****(10): p. 2100091.*

39. *Durao, A., et al. On the Quantization of Recurrent Neural Networks for Smiles Generation. in ICASSP 2023 - 2023 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). 2023.*
40. *Hu, H., et al., Machine learning for the prediction of safe and biologically active organophosphorus molecules. 2023.*
41. *Nowak, D., R.A. Bachorz, and M. Hoffmann Neural Networks in the Design of Molecules with Affinity to Selected Protein Domains. International Journal of Molecular Sciences, 2023.* ***24****,*
42. *Svensson, H.G., et al., Utilizing Reinforcement Learning for de novo Drug Design. arXiv preprint arXiv:2303.17615, 2023.*
43. *Popova, M., O. Isayev, and A. Tropsha, Deep reinforcement learning for de novo drug design. Science Advances.* ***4****(7): p. eaap7885.*
44. *Pereira, T., et al., Diversity oriented Deep Reinforcement Learning for targeted molecule generation. Journal of Cheminformatics, 2021.* ***13****(1): p. 21.*
45. *Atance, S.R., et al., De Novo Drug Design Using Reinforcement Learning with Graph-Based Deep Generative Models. Journal of Chemical Information and Modeling, 2022.* ***62****(20): p. 4863-4872.*
46. *Mazuz, E., et al., Molecule generation using transformers and policy gradient reinforcement learning. Scientific Reports, 2023.* ***13****(1): p. 8799.*
47. *Tan, Y., et al., DRlinker: Deep Reinforcement Learning for Optimization in Fragment Linking Design. Journal of Chemical Information and Modeling, 2022.* ***62****(23): p. 5907-5917.*
48. *Ishitani, R., T. Kataoka, and K. Rikimaru, Molecular Design Method Using a Reversible Tree Representation of Chemical Compounds and Deep Reinforcement Learning. Journal of Chemical Information and Modeling, 2022.* ***62****(17): p. 4032-4048.*
49. *Pereira, T.O., et al., End-to-end Deep Reinforcement Learning for Targeted Drug Generation, in Proceedings of the 2020 4th International Conference on Computational Biology and Bioinformatics. 2021, Association for Computing Machinery: Bali Island, Indonesia. p. 7–13.*
50. *Ranjan, A., et al., Molecule generation toward target protein (SARS-CoV-2) using reinforcement learning-based graph neural network via knowledge graph. Network Modeling Analysis in Health Informatics and Bioinformatics, 2023.* ***12****(1): p. 13.*
51. *Goel, M., et al., MoleGuLAR: Molecule Generation Using Reinforcement Learning with Alternating Rewards. Journal of Chemical Information and Modeling, 2021.* ***61****(12): p. 5815-5826.*
52. *Zhou, Z., et al., Optimization of Molecules via Deep Reinforcement Learning. Scientific Reports, 2019.* ***9****(1): p. 10752.*
53. *Tong, X., et al., Generative models for de novo drug design. Journal of Medicinal Chemistry, 2021.* ***64****(19): p. 14011-14027.*
54. *Lee, M. and K. Min, MGCVAE: Multi-Objective Inverse Design via Molecular Graph Conditional Variational Autoencoder. Journal of Chemical Information and Modeling, 2022.* ***62****(12): p. 2943-2950.*
55. *Boitreaud, J., et al., OptiMol: Optimization of Binding Affinities in Chemical Space for Drug Discovery. Journal of Chemical Information and Modeling, 2020.* ***60****(12): p. 5658-5666.*
56. *Liao, Z., et al., Sc2Mol: a scaffold-based two-step molecule generator with variational autoencoder and transformer. Bioinformatics, 2023.* ***39****(1): p. btac814.*
57. *Lin, Z., et al., MoVAE: A Variational AutoEncoder for Molecular Graph Generation, in Proceedings of the 2023 SIAM International Conference on Data Mining (SDM). 2023, Society for Industrial and Applied Mathematics. p. 514-522.*
58. *Bhadwal, A.S. and K. Kumar. GVA: Gated Variational Autoencoder for de novo molecule generation. in 2022 IEEE 9th Uttar Pradesh Section International Conference on Electrical, Electronics and Computer Engineering (UPCON). 2022.*
59. Harada, Y., et al., *Molecular Design Learned from the Natural Product Porphyra-334: Molecular Generation via Chemical Variational Autoencoder versus Database Mining via Similarity Search, A Comparative Study.* ACS Omega, 2022. **7**(10): p. 8581-8590.

60. Ma, C. and X. Zhang, *GF-VAE: A Flow-based Variational Autoencoder for Molecule Generation*, in *Proceedings of the 30th ACM International Conference on Information & Knowledge Management*. 2021, Association for Computing Machinery: Virtual Event, Queensland, Australia. p. 1181–1190.
61. Krishnan, K., et al. *Interpretable Machine Learning Models for Molecular Design of Tyrosine Kinase Inhibitors Using Variational Autoencoders and Perturbation-Based Approach of Chemical Space Exploration*. International Journal of Molecular Sciences, 2022. **23**,
62. Kim, H., et al., *Predicting chemical structure using reinforcement learning with a stack-augmented conditional variational autoencoder.* Journal of Cheminformatics, 2022. **14**(1): p. 83.
63. Pravalphruekul, N., et al., *De Novo Design of Molecules with Multiaction Potential from Differential Gene Expression using Variational Autoencoder.* Journal of Chemical Information and Modeling, 2023. **63**(13): p. 3999-4011.
64. Blanco-Gonzalez, A., et al., *The role of ai in drug discovery: challenges, opportunities, and strategies.* Pharmaceuticals, 2023. **16**(6): p. 891.
65. Moingeon, P., M. Kuenemann, and M. Guedj, *Artificial intelligence-enhanced drug design and development: Toward a computational precision medicine.* Drug discovery today, 2022. **27**(1): p. 215-222.
66. Askr, H., et al., *Deep learning in drug discovery: an integrative review and future challenges.* Artificial Intelligence Review, 2023. **56**(7): p. 5975-6037.
67. Sarkar, C., et al., *Artificial intelligence and machine learning technology driven modern drug discovery and development.* International Journal of Molecular Sciences, 2023. **24**(3): p. 2026.
68. Chen, W., et al., *Artificial intelligence for drug discovery: Resources, methods, and applications.* Molecular Therapy-Nucleic Acids, 2023.
69. Gill, S.S., et al., *AI for next generation computing: Emerging trends and future directions.* Internet of Things, 2022. **19**: p. 100514.
70. Sebastian, A.M. and D. Peter, *Artificial intelligence in cancer research: trends, challenges and future directions.* Life, 2022. **12**(12): p. 1991.