How to Cite:
Alanazi, A. A. A., Al Fahad, Abdulrahman I. A., Almorshed, A. S. A., Alrbian, Abdullah A. M., Alnughaymishi, A. A. S., Al-Mutairi, N. H. B., Alajmi, A. A., & Al Otaibi, S. G. (2022). Artificial intelligence in drug discovery: Current applications and future directions. *International Journal of Health Sciences*, *6*(S10), 2011–2040. https://doi.org/10.53730/ijhs.v6nS10.15290

Artificial intelligence in drug discovery: Current applications and future directions

Ahmad Asri Awad Alanazi
Health Informatics technician, National Guard Health Affairs

Abdulrahman Ibrahim Abdullah Al Fahad
Health Informatics technician, National Guard Health Affairs

Abdullah Saleh Abdullah Almorshed
Pharmacy Technician, National Guard Health Affairs

Abdullah Abdulrahman Mohammed Alrbian
Pharmacy Technician, National Guard Health Affairs

Ahmed Abdulkarim Sulaiman Alnughaymishi
Health Informatics technician, National Guard Health Affairs

Nawal Haif Bajad Al-Mutairi
National Guard Health Affairs

Abdulrahman Amer Alajmi
National Guard Health Affairs

Mr. Sanytan Ghazy Al Otaibi
National Guard Health Affairs

***Abstract*---Background:** The drug discovery process is complex, time-consuming, and costly, traditionally relying on trial-and-error approaches. The integration of artificial intelligence (AI) and machine learning (ML) has emerged as a transformative solution, enhancing efficiency and precision in identifying potential drug candidates. **Aim:** This review aims to explore the current applications of AI in drug discovery, highlight the AI tools utilized in the process, and discuss the associated challenges. **Methods:** A comprehensive literature review was conducted, focusing on peer-reviewed articles, clinical studies, and case reports that detail the application of AI and ML in various phases of drug discovery, including target identification, lead

International Journal of Health Sciences E-ISSN 2550-696X © 2022.
***Corresponding email**: akk29337@gmail.com*
Manuscript submitted: 01 Jan 2022, Manuscript revised: 09 Jan 2022, Accepted for publication: 15 Jan 2022

optimization, and preclinical evaluation. **Results:** The review identifies several AI applications, such as predictive modeling, molecular design, and virtual screening, which significantly expedite the discovery process. Tools such as deep learning, natural language processing, and reinforcement learning have been instrumental in analyzing large datasets and predicting drug interactions. However, challenges remain, including data integration issues, skill gaps among professionals, and skepticism regarding AI outcomes. **Conclusion:** AI has the potential to revolutionize drug discovery by streamlining processes and improving accuracy. However, addressing the challenges related to data management, workforce training, and trust in AI systems is essential for realizing its full potential in the pharmaceutical industry.

***Keywords*---**Artificial Intelligence, Drug Discovery, Machine Learning, Predictive Modeling, Pharmaceutical Industry.

Introduction

Drug discovery is widely recognized as an arduous, costly endeavor with a historically low success rate. Typically, it requires an investment of 2.6 billion USD [1] and spans over a decade. Furthermore, the probability of progressing a drug from Phase I clinical trials to market is discouragingly low, with a success rate below 10% [2]. Over the last ten years, the landscape of drug discovery has been undergoing significant shifts, largely fueled by rapid advances in artificial intelligence (AI) [3–7]. AI is now frequently utilized in various drug discovery applications, such as virtual screening [8], de novo drug design [9], retrosynthesis and reaction prediction [10], and de novo protein design [11]. These applications can generally be classified into predictive and generative tasks. A range of AI techniques powers these tasks, with model architectures advancing from traditional machine learning approaches to deep neural networks, including convolutional neural networks, recurrent neural networks, graph neural networks, and transformers. Additionally, learning paradigms have expanded from supervised learning to self-supervised and reinforcement learning models.

Innovation remains essential to drive progress across all fields, particularly within medicine and pharmaceuticals, where the primary objective is the discovery and development of compounds to alleviate human suffering. For decades, pharmaceutical production has been governed by rigorous regulatory standards that ensure quality control through testing raw materials, in-process substances, and final product attributes, along with batch-specific procedures and controlled conditions [12]. However, the drug and biopharmaceutical sectors have historically offered limited advancements in terms of innovative machinery or groundbreaking principles. Consequently, these industries face an urgent demand for technological innovations to streamline drug development for safe human consumption. Manufacturing complex drugs at a commercial scale and integrating them into standard therapeutic practices remain challenging due to technological constraints [12]. Current pharmaceutical prescribing approaches are often generic, following a “one size fits all” model. However, advancements in

genomics and diagnostics now allow for personalized medicines crafted to match a patient's biological profile. The industry's current technological capabilities are inadequate to meet the demands of personalized medicine, indicating a need for novel manufacturing solutions and customizable equipment within the pharmaceutical sector [12].

The adoption of AI is rapidly transforming clinical research and training. Healthcare professionals can engage in this technological advancement to ensure AI's potential for enhancing medical care is fully realized [13]. In the pharmaceutical sector, AI applications encompass four primary areas: assessing disease severity and predicting treatment outcomes for individual patients; preventing or resolving treatment-related complications; supporting treatment procedures; and analyzing the purpose of specific instruments or chemicals, leading to novel uses for improved safety and efficacy. In addition to these uses, AI plays a pivotal role in the management and analysis of big data [14]. Big data, a paradigm focusing on vast datasets combined with advanced analytics, has opened up new research opportunities [15], [16]. As traditional data storage approaches become inadequate for handling the industry's growing data volume, big data enables deeper research via data mining. A three-step data management process—data extraction and consolidation, formatting for uniformity, and analysis using diverse platforms—enhances decision-making regarding compound selection and manufacturing processes [16]. With the recent surge in AI-driven technologies, coupled with the data influx available for insightful analysis, the pharmaceutical industry is actively integrating these innovations to address minor but critical challenges in drug development. This prompts the need for a comprehensive exploration of AI, machine learning, and big data's role in advancing drug discovery, summarizing recent technological advancements, studies, and future implications as the industry integrates the latest AI advancements into practice.

Drug Discovery Process:

When there are no treatments for a given condition or when the pharmaceuticals that are currently on the market are not working well because of their high toxicity or poor efficacy, drug discovery [16] is the solution. Target identification and validation are the first phases in the process, which starts with developing a hypothesis that the activation or inhibition of a particular target—such as an enzyme, receptor, or ion channel—will result in therapeutic benefits. After a target has been chosen, extensive assays are carried out to find early hits and turn them into leads, or possible drug candidates. Lead optimization, hit discovery, and the hit-to-lead transition are all steps in this process. Following their development, prospects move on to preclinical and clinical testing before, if successful, being introduced as medicinal products. High-throughput screening (HTS) [17,18] was developed in the 1980s as a high-efficiency hit detection technique that relies on automation and large chemical libraries to improve the effectiveness of small-molecule drug discovery. One significant result of HTS has been the creation of extensive structure–activity relationship (SAR) datasets, which are used to fill chemical databases like ZINC [20] and PubChem [19]. As in vitro and in vivo testing progresses, computational methods—collectively referred to as virtual screening (VS)—have been developed to analyze large chemical libraries in order

to find molecules that are probably biologically active [21]. In order to increase the likelihood of finding active compounds, VS uses computational techniques to identify active molecules either based on the target structure (structure-based VS) or previously identified active ligands (ligand-based VS) [22]. The idea of active chemicals, which are often categorized into agonists and antagonists based on their mechanism of action (MoA), is consistent with the fundamental theory that a disease can be treated by activating or inhibiting a target [23]. An antagonist binds to prevent a biological reaction, while an agonist binds to activate a target and provide a response akin to that of the endogenous ligand. Additionally, antagonists come in competitive and non-competitive varieties, and agonists can be further classified as partial, inverse, or biased agonists. These molecules' efficacy (effect magnitude) and affinity (binding capacity) are measured by a variety of assays; potency, which is inversely correlated with affinity, indicates the amount required to produce a particular effect level, while affinity indicates how much of a molecule binds to a target at a given concentration. Efficacy is a measure of the effect's strength, such as 60% inhibition of an enzyme.

However, binding selectivity is also essential, thus a molecule is not necessarily a great therapeutic candidate only because it satisfies activity requirements [24]. A chemical may frequently attach to more than one target, which could result in unexpected side effects. High selectivity is therefore equally important. Drug candidates must meet a number of requirements [25], including pharmacokinetic (such as absorption, distribution, metabolism, and excretion), physicochemical (such as solubility, acid-base characteristics, lipophilicity, and permeability), and pharmacodynamic (such as activity, selectivity) traits. Factors like the Quantitative Estimate of Drug-likeness (QED) and the Synthetic Accessibility Score (SAS) are also assessed during drug production. While QED (rated between 0 and 1) indicates the probability that a molecule is a good candidate for a medicine, SAS is a heuristic evaluation (from 1 to 10) that estimates the synthesis difficulty based on molecular fragment contributions. Drug discovery fundamentally involves multi-objective optimization [26]. Quantitative structure–activity relationship (QSAR) modeling is the general term for the process of creating predictive models for each pertinent property in order to map the molecular structure to the property value through classification or regression [27]. One intriguing area of QSAR is de novo drug design, which involves applying these models conversely to find structural characteristics that improve desirable qualities and lead the creation of new medications from the ground up.

Drug design aims to investigate the large chemical universe, an enormous area that includes all potential tiny molecules, estimated to vary in size from 10^30 to 10^60 [29, 30]. This goes beyond screening existing chemical libraries [28]. Repeated organic synthesis and property evaluation are part of the iterative design-make-test-analyze (DMTA) cycle used in drug design [3]. The late 1970s saw the emergence of quantitative drug design techniques [31], which were based on two main inquiries [32]: (i) "Can molecular properties be deduced from molecular structures?" as well as (ii) "Which structural characteristics influence specific molecular properties?" Whereas QSAR answers the latter question, VS is based on the former. According to this perspective, drug design goes beyond VS to encompass the creation of novel molecules as well as the prediction of molecular

properties, which are the main goals of today's AI-enabled drug discovery methodologies.

Advanced Technologies in Drug Discovery Process:

To improve drug research and production efficiency, the pharmaceutical sector incorporates a variety of cutting-edge technology. Among these, Bayer (Leverkusen, Germany) developed the in silico ADMET (Absorption, Distribution, Metabolism, Excretion, and Toxicity) platform, which makes it easier to predict the physicochemical and pharmacokinetic features of novel drugs [33]. This platform can function using two distinct approaches: the first involves assessing how compounds interact with particular proteins, which calls for high-definition 3D imaging of each protein because of its direct impact on ADMET factors like cost-effectiveness, safety, and cognition [33]. Its dependence on individual proteins with distinct functions is a drawback, though. The second approach, on the other hand, collects secondary data from a variety of chemicals in order to create intricate hybrid models that employ AI and machine learning to forecast ADMET attributes (Bayer, Leverkusen, Germany) [33]. Due to the requirement for constant data maintenance, this method lengthens project durations even if it necessitates significant manual oversight and exact adjustments to avoid expensive errors [33]. Because of its permanence, decentralization, transparency, and traceability, blockchain technology—a system that arranges data chronologically in an unchangeable chain—also shows promise in the pharmaceutical industry [34]. Each feature helps pharmaceutical companies track data effectively: decentralization allows many entities to handle the data, while permanence secures the data. While traceability offers a thorough history of recorded data via timestamps, transparency enables users to access all stored data [34]. Although blockchain applications have demonstrated advantages in the supply chain process and the effectiveness of medication trials, smaller businesses and lower-income regions are unable to utilize the technology due to its high cost [34]. Last but not least, 3D printing offers a promising development. New techniques that use blue light to improve accuracy and maximize material impedance transfer enable real-time, incredibly exact reproductions [35-36]. Due to its growing accessibility and usefulness, these developments not only make on-site, on-demand production easier, but they also expand the possible uses of 3D printing across industries [35-36]. These new technologies nevertheless face operational and financial obstacles in spite of these developments, which suggests that further innovation is required to increase the pharmaceutical industry's affordability, effectiveness, and user accessibility.

Drug Discovery and AI:

Since the early 2000s, AI has had a major impact on drug development. Machine learning techniques, including random forest (RF), have been used for quantitative structure-activity relationship (QSAR) modeling and virtual screening (VS) [37-38]. The deep learning era began in 2012 with the release of AlexNet [39-40], and it quickly showed promise in this area. For example, deep neural networks (DNNs) fared better than RF models in predicting chemical activity during the 2012 Merck Kaggle competition [37]. Deep learning applications in chemistry have grown, and AI developments in computer vision and natural

language processing have further aided in drug discovery [41-42]. Notable innovations include the development of halicin, an antibiotic contender against resistant bacteria, by MIT researchers in 2020 [44] and the quick discovery of strong inhibitors for discoidin domain receptor 1 (DDR1) by Insilico Medicine in 21 days (2019) [43].

From finding and validating targets to assessing medication response, artificial intelligence is involved in many phases of drug discovery [6]. Lead identification—more especially, molecule creation and molecular property prediction—is the main topic of this survey. In order to enable tasks such as drug-target interaction (DTI), toxicity, and drug-induced liver damage (DILI) prediction, molecular property prediction seeks to ascertain a molecule's attributes based on its structure or learnt representation [45-47]. The two objectives of molecule creation—realistic generation, which follows chemical principles, and goal-directed generation, which creates compounds with desired properties—are crucial to drug design [48-49]. As many have pointed out, despite the advancements, difficulties still exist in AI-driven drug discovery [50-53]. "A method cannot save an unsuitable representation which cannot remedy irrelevant data for an ill-thought-out question," according to Bender et al. [51-52]. It is essential to take a comprehensive approach when evaluating elements like hypotheses, data quality, representations, models, and learning paradigms in AI applications for drug discovery in order to control expectations and avoid overhype.

Machine Learning and Drug Discovery:

The pharmaceutical business is seeing a sharp increase in the use of machine learning (ML), especially in drug discovery, where it has enhanced industry-wide procedures. The growing number of businesses that make machine learning a key component of their business strategy is proof of this progress. To improve drug research and development (R&D), big pharmaceutical companies have also looked into machine learning (ML) approaches [54]. Future developments in the field of drug discovery must include these methods due to the broad potential of machine learning. Utilizing high-throughput screening methods, which have the potential to drastically cut down on the time and expense of drug discovery, is one of the main goals. Live animal testing may even be less necessary or perhaps eliminated because of ML [55]. Research demonstrates ML's usefulness in this field and supports it as a useful tool. Specific chemical and biological data are needed to improve machine learning applications in drug discovery. By offering insights derived from quantifiable medicinal characteristics, such as cellular toxicity, cell structure heterogeneity, animal model efficacy, on-target activity, pharmacokinetic endpoints, microsomal stability, and cytochrome P450 (CYP) inhibition values, such data can improve the precision and performance of machine learning (ML) systems [55]. The simplified cycle of machine learning's use in the pharmaceutical sector, specifically in development and manufacturing. Here, a preset algorithm processes a variety of data inputs, including orthogonal and application domain data. With the use of model selections and orthogonal calculations, this algorithm generates results and iterative enhancements, continuously improving the process until the finished output is produced [56].

Margulis and colleagues [57] carried out a practical demonstration of machine learning's use in drug discovery, examining the technology's capacity to detect very bitter compounds at an early stage of drug development. The objective was to determine if an ML algorithm could predict the bitterness of medicinal molecules in place of animal testing. Interestingly, 80% of the bitter compounds found matched the findings of a brief access taste aversion (BATA) test, indicating the effectiveness of ML. Additionally, the study offered fresh perspectives by demonstrating that toxicity and bitterness are not necessarily related. Further studies on ML's function in G-protein coupled receptor (GPCR) ligand recognition—which is essential for drug discovery—were carried out by Raschka et al. [58-59]. Finding out if machine learning (ML) might take the place of conventional technologies in Sea Lamprey Receptor 1 (SLOR1) signal inhibition tests was the aim. The findings of the ML algorithm showed that it may be a good substitute for recognizing chemical features because they nearly matched baseline performance. In order to determine whether an ML algorithm could efficiently rank docking receptors, and their corresponding ligands, Pereira and colleagues carried out an additional investigation. The results showed that the ML algorithm performed better than conventional ranking techniques, indicating that ML has potential for creating molecular databases, identifying pharmacokinetic properties, and classifying compounds.

Another noteworthy work achieved higher accuracy than non-linear approaches by combining linear description analysis with a variety of descriptors to predict antifungal and antibacterial capabilities across diverse fungus and bacteria. This demonstrates the usefulness of machine learning (ML) in the drug development process by showing that it can help find unique drug features. In a separate study, deep learning and machine learning algorithms were used to categorize medications according to how they affected genes. The results showed a notable improvement in accuracy, indicating that deep learning models can be incorporated into machine learning frameworks to enhance classification accuracy and facilitate efficient drug classification according to pharmacodynamic and pharmacokinetic characteristics. Two further studies demonstrate the usefulness of particular ML techniques. Reinforcement learning (RL) was used by Rantanen and Khinast [60] to speed up the process of finding potentially useful medications. Using a different strategy, Turki and Taguchi [61] predicted how multiple myeloma patients will react to an existing medication by using transfer learning. Zhavoronkov and Mamoshina showed improved predictive accuracy with transfer learning, while Turki and Taguchi discovered that RL reduced drug discovery durations to 46 days. These researchers concluded that these methods could be further improved by expanding their applications and improving algorithms [62-66].

Several studies have utilized machine learning (ML) approaches in drug discovery, achieving notable accuracy levels across different applications. Traditional reinforcement learning has been applied in new drug development by integrating multiple ML techniques to generate new molecules, achieving high accuracy with 95% of molecules deemed feasible for development [62]. Transfer learning has also been used to model biological responses, specifically for anticancer drugs, through regression-based techniques, which proved highly accurate in emulating biological processes [63]. Additionally, multitask learning has been leveraged to

support drug development and testing by monitoring genetic and medicinal data, accurately tracking the pathways of drug molecules [64]. In drug–target interactions, multitask analysis combined various ML algorithms to analyze how drugs interact with their targets, yielding precise insights [65]. Lastly, multitask learning techniques have been applied in post-manufacturing drug reviews, where bulk data analysis using these algorithms enabled the efficient processing of 4,200 reviews in a short time frame, demonstrating a high level of accuracy [66].

AI and Drug Discovery:

With advancements in computational power and AI techniques, the drug discovery and development process could undergo substantial reform. The pharmaceutical industry is currently challenged by declining efficiency in drug improvement programs and rising R&D costs [67]. In recent years, digital information within the industry has expanded significantly; however, effectively gathering, analyzing, and applying this data to address complex clinical issues remains challenging. AI is well-suited to handle vast data volumes with enhanced automation capabilities [68]. By incorporating machine learning algorithms, AI has the potential to boost efficiency and productivity in drug discovery. This section explores AI's primary applications to enhance the effectiveness of the drug discovery cycle.

This process is divided into four stages: drug design, polypharmacology, drug repurposing, and drug screening [69]. A central use of AI is in predicting drug properties, which could reduce the need for clinical trials and live study participants, providing both financial and ethical benefits. Studies reviewed here underscore AI’s potential to improve efficiency, accuracy, and productivity in drug discovery. In one study, Cui and Zhu [70] utilized a neural network, ResNet, to predict drugs' physicochemical properties, including solubility and partition coefficient. Results showed that ResNet more accurately predicted molecular solubility compared to non-AI models, demonstrating AI’s ability to enhance efficiency in drug development. Another study by Lusci et al. [71] used a recursive neural network to predict the solubility of chemical and biological molecules in water, confirming AI’s improved accuracy over conventional techniques and supporting its role in drug discovery. Polykovskiy and colleagues [72] explored AI’s role in predicting synthesized molecules' activity to increase drug screening reliability. Using an adversarial auto-encoder, they generated compounds with unique features, complementing results from a Recurrent Neural Network-based model and unveiling new insights into molecule-target interactions. This study highlighted AI’s dual capacity to enhance accuracy and enable novel discoveries.

Daynac et al. [73] further examined AI’s efficiency by predicting molecules' antimicrobial properties, achieving over 70% accuracy with a 10 mm error margin. The neural network also handled multiple molecules simultaneously, reducing processing time. Similarly, Pu et al. [74] used eToxPred, an AI model, to predict toxicity levels in synthetic and biological compounds, achieving a 72% accuracy rate and a minimal error margin of 4%, which indicates its potential as a clinical trial alternative. Two studies, by Kadura et al. [75] and Maram and Hamdy [76], applied different AI techniques to anticancer drug screening and biological activity prediction. Kadura’s study utilized adversarial networks to

screen approximately 72 million anticancer molecules, showcasing AI's efficiency. Maram and Hamdy employed various AI models, achieving a 71.9% accuracy rate, which surpassed baseline readings (RF: 62.6% and SVM: 66.0%), with a low error margin. Together, these studies illustrate AI's diverse applications and value across the pharmaceutical landscape [75-79].

Recent studies illustrate the innovative use of artificial intelligence (AI) in drug discovery, employing various machine learning techniques to enhance accuracy and efficiency. Deep learning, for example, has been applied in drug screening by leveraging protein-ligand interactions as input values for DeepTox to predict drug toxicity with high accuracy [75]. Neural networks have similarly advanced drug design, where deep learning neural networks are used to predict protein structures, achieving impressive accuracy [76]. Reinforcement learning also plays a key role in drug screening, applying machine learning techniques to identify inhibitor molecules for specific proteins, such as DDR1, with reliable precision [77]. Additionally, neural networks and deep learning methods predict interactions between drugs and their targets in drug design, with results indicating high accuracy [78]. Another notable application integrates neural networks within a neural computer framework to design small organic molecules, achieving exceptional accuracy in new molecule design [79]. These studies collectively underscore AI's transformative role in optimizing various aspects of drug discovery.

Successful Applications of AI in Drug Discovery:

Many evaluations that highlight the growing role of artificial intelligence (AI) in drug discovery have surfaced in recent years [80-81] Therefore, the purpose of this review is to shed light on a number of excellent research that show the real-world effects of AI in small-molecule drug development.

Drug Target Structural Enablement and Binding Site Comparisons:

The availability of atomic-resolution structural information about small molecules interacting with drug targets opens up possibilities for fragment-based drug discovery (FBDD), ligand optimization (SBDD), and structure-guided hit identification (SVA). These structural data also provide information about the mechanisms of resistance, modes of action, allosteric pocket discovery, selectivity, and ligandability evaluation for new drug targets [82-83]. Only around 35% of the human proteome is currently covered structurally, despite major advances in single-particle cryo-electron microscopy (cryo-EM), nuclear magnetic resonance (NMR) spectroscopy, and X-ray crystallography. In a multidomain protein, its coverage is frequently limited to a single structural domain. As a result, the number of known protein sequences and the empirically resolved structures continue to differ significantly [84]. Notably, databases like the Protein Data Bank (PDB) have inadequate structural representation for pharmacologically relevant protein target families, including ion channels and G-protein-coupled receptors (GPCRs) [85]. Computational techniques for protein 3D structure prediction are a substitute strategy when experimental structures are unavailable. By predicting the three-dimensional conformation of an unknown (target) protein using the experimental structure of a homologous (template) protein and its amino acid

sequence, homology modeling has long been used to close the sequence-structure gap. According to research, SBDD may often use homology-modeled structures with a sequence identity as low as 30% [86]. Accurately predicting the structure of proteins without a homologous structure is still difficult, but recent developments in deep learning (DL)-based approaches and the use of coevolution data in modeling have brought protein structure prediction back to life [87]. Even without a template, DL algorithms including generative adversarial networks, variational autoencoders (VAEs), recurrent neural networks (RNNs), and convolutional neural networks (CNNs) have demonstrated increased performance in protein structure prediction [88].

With the impressive performance of a deep convolutional residual network (ResNet)-based algorithm called AlphaFold2 at the CASP14 competition, the use of DL approaches for protein structure prediction attracted a lot of attention [89]. AlphaFold2, created by DeepMind Technologies [90], predicted the distribution of distances between pairs of amino acids and the torsion angles between chemical bonds connecting those amino acids within a protein using a deep neural network (DNN) architecture trained on 170,000 protein structures from the PDB. Additionally, it uses an end-to-end folding technique for structure prediction and integrates evolutionary information obtained from numerous sequence alignments. AlphaFold2's architectural framework and methodology were just released [91]. The Global Distance Test (GDT) metric, which measures the correspondence between residues in the model and the experimental structure, is used by CASP to evaluate the precision of structure predictions. The prediction accuracy is equivalent to experimental approaches when the GDT score is 90; AlphaFold2 obtained a median score of 92.4 GDT for all targets. The CASP14 competition results showed that DL techniques can attain impressive accuracy levels that are on par with experimental structures. With structural coverage for 98.5% of the human proteome, DeepMind and EMBL-EBI have made the 3D structures predicted by AlphaFold2 publicly available to the scientific community [92]. Inspired by AlphaFold2's success, a research team under David Baker also created RoseTTAFold, a three-track neural network program [93].

It is still too soon to declare that AI has solved the protein-folding conundrum or that its impact on drug discovery will be revolutionary, even though these developments mark progress in protein structure prediction. More than 170,000 protein structures from the PDB were used to train AlphaFold2, and the quality of the data used to train any learning model is intrinsically related to its effectiveness. Although the actual number of folds found in nature is estimated to be between 4,000 and 10,000, current estimations based on the SCOP version 2 database show that 1,388 folds are recorded in the PDB [94-95]. As a result, the PDB is missing many new folds, topologies, and architectures and has significant redundancy in both sequence and protein families. AI also has difficulties in predicting the structure of membrane proteins, multidomain proteins, and multimeric protein complexes. However, DeepMind's technology has the potential to advance structural biology, speed up drug discovery, and facilitate de novo protein creation. One useful exercise in structure-based medication design is the comparative examination of protein-binding pockets at the structural genomics scale [96]. It provides information that improves selectivity comprehension, forecasts off-target liabilities, guides efforts to repurpose drugs, and aids in the

annotation of protein functions. Traditional pocket comparison techniques calculate sequence-free similarities across binding sites using representations like graph theory, geometric hashing, typed triangles, spherical harmonics, and physicochemical properties of binding site atoms. These intuition-based feature extraction techniques, however, are usually not scalable across several binding sites and may introduce human biases.

The construction of strong voxel-based feature representations that can capture chemical features and translate binding locations into descriptor vectors has been made easier by the introduction of deep neural network (DNN) methods. An example of this is DeeplyTough [97], which uses 3D steerable CNNs to compare binding locations without alignment. To do this, 3D representations of protein pockets are encoded into descriptor vectors, which can then be used to compute pairwise Euclidean distances in order to measure pocket similarity. The TOUGH-M1 dataset, a non-redundant and representative set of small-molecule binding pockets with over a million data points, served as the training set for DeeplyTough. A negative subset of distinct proteins binding chemically dissimilar ligands and a positive subset of different proteins binding chemically comparable ligands are both included. DeeplyTough's performance was assessed on two separately created datasets (Vertex and ProSPECCTs), showing competitive outcomes when compared to current approaches and a notable reduction in runtime.

Using AI to Enhance Virtual Screening:

A complimentary and economical alternative to experimental techniques such as high-throughput screening for hit detection is virtual screening (VS), a computer approach [98-101]. Instead of physically screening every molecule in a collection, VS uses computer methods to rank a subset of compounds for initial test evaluation. Traditional VS approaches face considerable hurdles due to the rapidly growing amount of "make-on-demand" screening libraries and the increasing number of high-value, difficult pharmacological targets discovered by functional genomics. As a result, there is a lot of interest in drug development in AI techniques that improve VS methodologies and enable effective chemical space exploration for hit identification.

Ligand-Oriented Virtual Inspection:

The goal of ligand-based virtual screening (LBVS) methods is to use molecular similarity principles to find active chemicals from a chemical library. These methods include predictive modeling, shape matching, pharmacophore screening, and similarity searching. The traditional quantitative structure-activity relationship (QSAR) paradigm is expanded upon by predictive modeling for VS. Conventional QSAR builds explanatory models that measure structure-activity relationship (SAR) trends in the past by using statistical data modeling techniques on a congeneric series. New opportunities for QSAR modeling as a VS approach have been made possible by the availability of extensive chemogenomics data (such as PubChem's BioAssay (64) and the ChEMBL database [102] as well as developments in machine learning (ML) and deep learning (DL) algorithms that can handle big datasets. Consequently, there are many documented effective uses

of QSAR-based VS processes for hit identification. Zhang et al., for example, described how an ML-based QSAR methodology for VS was successfully implemented and resulted in the discovery of novel antimalarial drugs [103]. The researchers developed a binary classifier model (active or inactive) on 3,133 chemicals with known antimalarial properties using two machine learning algorithms, support vector machines (SVM) and k-nearest neighbors (kNN). Following VS against the ChemBridge database using the QSAR models, 174 compounds were chosen for additional screening in Plasmodium falciparum growth inhibition and cellular tests. 25 of the chosen compounds demonstrated action, according to experimental validation, with the most potent hit having an EC50 value of 95.6 nM and a hit rate of 14.2%. The application of ML and DL-based QSAR procedures as prospective VS tools has since been shown in a number of research [104-110].

The use of web-based cheminformatics workbenches to automate and optimize ML- and DL-based QSAR operations for VS has increased within the past ten years. A user-friendly open-source web server called DeepScreening [111] was created by Liu et al. to allow users to construct and verify recurrent neural network (RNN) models using either user-provided datasets for VS or ChEMBL bioactivity data. Using carefully selected bioactivity data from ChEMBL 24, DeepScreening additionally provides prebuilt deep neural network (DNN) models for 1,251 targets. Both professionals and nonexperts can execute VS against certain targets of interest because to this user-friendly interface and the prebuilt QSAR models that are available. Another open-source web server that uses machine learning (ML) to create categorical QSAR models from PubChem data is DpubChem [112]. Despite the fact that chemogenomic databases such as PubChem and ChEMBL offer a wealth of bioactivity data for model construction, using these resources presents several difficulties. The existence of bioactivity data from several sources and an unbalanced ratio of active to inactive molecules for any particular target are major causes for concern. Comparing this imbalance to other virtual screening methods makes it more difficult to generalize QSAR-based workflows for VS when using public datasets.

Virtual Screening Based on Structure:

From hit identification to lead optimization (LO) and binding mode prediction, molecular docking is a popular computational technique in structure-based drug design (SBDD) and is essential to many structurally enabled drug discovery programs [113]. In the docking method, the bound ligand conformation (position prediction) within a binding site is predicted, and then the binding affinity (scoring) is estimated [114]. The use of AI techniques in SBDD has grown in popularity due to notable developments in DL. DL is capable of autonomously learning and extracting features from 3D structural data, in contrast to shallow-learning AI techniques that depend on feature engineering [115]. As a result, multilayer feature extraction is being utilized to extract and generalize structural characteristics from protein-ligand complexes using DL approaches, which are frequently used in image recognition. This advancement creates new opportunities for using AI techniques in binding mode prediction, binding affinity estimate, and structure-based virtual screening (SBVS). Pose prediction and scoring during docking using ML and DL algorithms has demonstrated improved

performance in terms of screening power (ability to separate binders from non-binders), docking power (ability to distinguish native poses from decoy poses), and scoring power (ability to rank order compounds based on binding affinities) [116].

The first structure-based application to predict binding affinity using a deep convolutional neural network (CNN) framework was AtomNet® [117]. It uses a 3D grid technique to encode each atom's surroundings in the binding site into voxelized feature vectors. It was trained using a ChEMBL dataset that included 2,000,000 decoys and 78,000 active compounds across 290 targets. Interestingly, benchmark experiments utilizing the DUD-E benchmark dataset showed remarkable performance, with AtomNet® beating traditional docking approaches by reaching an area under the curve (AUC) over 0.9 for 57.8% of the targets. A significant obstacle in SBDD is the accurate docking-based prediction of ligand binding modalities. The binding mode seen by experimental methods like X-ray crystallography is frequently thought to correspond to the lowest energy pose, even though binding free energy is a macroscopic observable involving a ratio of partition functions between bound and unbound states [118]. Therefore, the top-scoring docked pose is chosen as its projected binding mode using the majority of classical scoring functions parameterized to duplicate binding affinities. However, it is still very difficult to reliably estimate binding energy using scoring functions, which frequently results in "hard failures," when the predicted binding modes do not match native or near-native binding modes.

Ashtawy and Mahapatra [119] made the first attempt to develop a task-specific scoring system for binding pose prediction using machine learning. They showed how several machine learning methods that could translate physicochemical and structural information from protein-ligand complexes could successfully differentiate between decoy and native docked postures. When compared to traditional scoring functions, the most efficient task-specific machine learning scoring function demonstrated enhanced docking power (>14%). The root mean square deviation (RMSD) of a predicted pose in relation to its native binding posture can be predicted by a comparable method called DeepBSP [120], which makes use of a 3D-CNN. The CASF-2016 benchmark dataset was used to assess the model's performance after it was trained on a dataset including 11,925 native complexes and more than 165,000 Autodock Vina [121] docked decoy poses. According to their findings, using DeepBSP to score Autodock Vina-generated postures showed better docking power than the application's hybrid knowledge and empirical-based scoring mechanism. An acetylcholinesterase (AChE) inhibitor with an IC50 value of 280 nM (Ki = 173 nM) was discovered by Adeshina et al. using "vScreenML" for VS in a prospective research [122]. Based on the XGBoost framework, vScreenML uses a classifier technique to classify docked poses as either decoys or actives. The scientists credited its success to the special characteristics of the training dataset, D-COID, which consists of both decoy complexes produced by molecular modeling and natural complexes that are representative of drug-like molecules. The vScreenML score function's classification accuracy was greatly improved by adding decoy complexes to the training dataset.

Recently, there has been a growing interest in a new class of scoring functions that are based on ML and DL techniques [123]. ML-based scoring functions have

been developed using a variety of popular ML and DL architectures, such as SVM, random forest (RF), kNN, gradient-boosting decision trees (GBDT), and 3D deep CNNs. Pafnucy [124], Onion-Net [125], RFScore-v3 (86), NNscore2.0 [126], BgN(BsN)-Score (82), and ΔvinaRF [127] are notable instances. A 3D CNN technique is used by both Onion-Net and Pafnucy to feature protein-ligand interactions. In numerous benchmark tests, ML-based scoring systems have continuously surpassed classical scoring functions [128]. Because ML-based scoring functions do not have a preset functional form that approximates the underlying physics of molecule recognition, they are nonparametric in contrast to standard scoring functions. Rather, they are trained on experimental datasets like PDBbind [129] and Binding MOAD [130], which contain both protein-ligand structure and binding affinity data. This enables them to implicitly account for molecular interactions that are challenging to model explicitly. ML-based scoring functions can be included into a docking program to direct pose sampling or used to rescore docked poses produced by an external docking program. The reader is directed to a number of outstanding reviews on the topic, as a thorough examination of ML-based scoring functions is outside the purview of this review [131]. Additionally, a number of studies document the effective application of ML and DL-based scoring systems that resulted in the discovery of hits that were empirically validated during VS campaigns. Yu et al., for instance, effectively found several hits against AChE and β-secretase-1 (BACE-1) by integrating the ML scoring algorithm with the docking software Smina (94). The influence of different ML-based scoring functions on enhancing VS outcomes was highlighted in a recently published systematic review that provided a thorough overview of them. The use of machine learning (ML)-based scoring systems in docking software is still relatively new, but it has a lot of potential to speed up drug development.

Active Learning Docking

As the volume of make-on-demand compound libraries expands, the demand for computational tools that effectively traverse the chemical space during virtual screening (VS) increases correspondingly. The ongoing discourse regarding whether a larger or smaller library yields superior outcomes in VS persists; however, the docking of ultralarge libraries for hit identification is gaining traction. The vast magnitude of these libraries presents significant challenges for docking programs, which typically struggle with brute force docking exceeding 100 million molecules. The incorporation of an active learning algorithm with molecular docking presents an innovative solution for enhancing the screening capacity of ultralarge libraries. Generally, active learning docking initiates by docking a small representative sample of the entire library, subsequently utilizing the outcomes to train a machine learning (ML) model that predicts docking scores for the remaining compounds. The compounds with the highest scores from the ML model are then docked, and the model is updated with the newly acquired data. This iterative process continues until the ML model achieves convergence. Thus, active learning docking maintains the precision of brute force docking while facilitating the identification of high-scoring compounds from extensive libraries, achieving this in significantly less time than traditional docking methods. An exemplary implementation of active learning docking is 'Active Learning Glide,' which synergizes the docking program Glide with the ML algorithms provided by

the open-source framework DeepChem. Graff et al. illustrated that a molecular-pool-based active learning-guided docking methodology successfully retrieved most of the top-scoring compounds in a virtual library at a fraction of the computational expense associated with brute force docking [132].

Generative Chemistry

The utilization of computational techniques for compound ideation has a well-established history. Early structure-based de novo design methodologies entailed the automated and incremental construction of ligands within receptor binding sites. Software such as LUDI identifies potential interaction sites within the binding pocket and assembles molecules from a predefined set of organic fragments that sterically and electronically complement the protein binding pocket. Inverse quantitative structure–activity/property relationship (QSAR/QSPR) modeling represents another de novo molecular design strategy that aims to create molecules with specific activity or properties by inversely mapping molecular descriptors from pre-existing QSAR/QSPR models. Generally, addressing the inverse-QSAR challenge is intricate, as reconstructing molecular structures from the descriptor information provided by a forward QSAR/QSPR model can be quite complex. A notable limitation hindering the widespread adoption of these classical de novo design strategies has been the restricted synthesizability and suboptimal drug-like characteristics of the designed molecules [133-135].

In recent years, the adoption of artificial intelligence (AI)-based generative modeling algorithms for de novo molecular design has seen a substantial rise, as they can effectively mitigate the challenges faced by traditional de novo design techniques. Generative chemistry leverages modern AI-driven generative modeling tools to produce synthetically feasible compounds possessing drug-like attributes while aligning with desired target property profiles. Employing a data-centric approach, generative modeling algorithms discern the underlying nonlinear distributions among molecular structures, their biological activities, and physicochemical properties from extensive datasets to inform compound design. In essence, AI-assisted generative modeling platforms facilitate compound ideation, predictive analysis, and selection of compounds with favorable characteristics. Various deep learning (DL) architectures, including variational autoencoders (VAEs), generative adversarial networks (GANs), reinforcement learning (RL), and recurrent neural networks (RNNs), have been employed in de novo molecular design. Current generative modeling techniques can also be categorized based on the molecular featurization method used. While early generative modeling approaches primarily utilized fingerprints and SMILES strings to encode molecular structures, newer methodologies, such as molecular graphs and fragment-based techniques, are gaining prominence [136].

A study by Kadurin et al. established a proof-of-principle for the application of deep GANs in generative modeling. Additionally, Segler et al. provided an example demonstrating the efficacy of RNNs in generative modeling. The inaugural report detailing the successful use of RNNs with long short-term memory (LSTM) cells for de novo molecular design was published by Gupta et al. Their generative LSTM model was trained on the ChEMBL database to produce novel molecules capable

of modulating retinoid X receptors (RXRs) and peroxisome proliferator-activated receptors (PPARs). The fine-tuning phase involved training on a dataset comprising 25 RXR and PPAR modulators (both agonists and partial agonists). The on-target activity of the generated compounds was evaluated using an ML model, resulting in four out of the five top-ranked compounds selected for synthesis demonstrating activity in a cell-based assay.

Another generative modeling investigation that garnered considerable media attention was the utilization of the GENTRL model by Zhavoronkov et al., which led to the identification of potent discoidin domain receptor 1 (DDR1) inhibitors within a span of only 21 days. The molecule designed via the GENTRL approach is depicted alongside its parent molecule and other DDR1 inhibitors. The authors employed a semi-supervised training method for their generative model, utilizing an objective function that incentivizes synthetic feasibility, target activity, and novelty. Out of the 30,000 compounds proposed by the generative model, six were synthesized and tested. Four compounds exhibited activity in biochemical assays, and two were active in cell-based assays, with the most effective compound displaying an IC50 value nearing 10 nM in both assay types. Despite demonstrating the capacity of generative modeling to identify a nanomolar hit compound, questions arose concerning the novelty of the molecules. The leading molecule bore striking similarities to the marketed multi-tyrosine kinase inhibitor ponatinib. Furthermore, the selectivity profile of the compound against the broader kinome was not established, casting doubt on its clinical significance. The authors addressed these critiques, clarifying that the study aimed to showcase the potential of generative modeling technology rather than identify a clinical candidate compound.

Recently, Perron et al. reported what may be the inaugural successful application of generative modeling in addressing a molecular property optimization (MPO) problem. An LSTM generative model combined with an RL method was employed to design 150 compounds predicted to fulfill all defined lead optimization (LO) criteria. The training dataset consisted of 881 molecules, encompassing 11 sets of associated assay data, including on-target activity, off-target activity, and ADMET endpoints. None of the training dataset compounds satisfied all property and potency criteria. Twenty compounds generated by the model were shortlisted for synthesis based on various criteria; nine of these failed during synthesis, while the remaining eleven were synthesized and profiled, with one meeting all eleven LO criteria. Despite the growing popularity of generative chemistry, it is crucial to emphasize the need for rigorous validation of generative models. Evaluation methodologies for generative modeling should encompass distribution learning benchmarks, synthetic validity, novelty, compound quality, and goal-directed objectives within their assessment framework. Open-source standardized benchmarking platforms, such as Molecular Sets (MOSES) and GuccaMol, could provide a framework for benchmarking generative modeling approaches [137-138].

In Silico ADMET Prediction

The recognition that inadequate pharmacokinetics of drug candidates significantly contributed to clinical attrition in the late 1990s prompted a

transformative shift within the pharmaceutical sector. This period saw the emergence of numerous property-based drug-likeness criteria, including Lipinski's Rule of Five, along with various developability metrics designed to monitor compound properties throughout the lead optimization process. Additionally, the advent of miniaturized, high-throughput in vitro ADMET profiling assays enabled the simultaneous assessment of efficacy and ADMET properties during the early phases of drug discovery. In silico ADMET modeling aims to assist project teams in the design and selection of innovative compounds with enhanced ADMET properties while directing experimental resources toward the most promising candidates, thereby minimizing the overall number of compounds requiring synthesis and profiling. Over the years, pharmaceutical companies have integrated numerous global in silico ADMET models into their discovery workflows.

Early endeavors in ADMET modeling predominantly employed linear regression techniques, such as those utilized by Hansch and Free-Wilson analysis. However, with the evolution of ML algorithms and the availability of extensive homogeneous ADMET datasets, silico ADMET modeling has transitioned toward ML-based predictive models developed using Bayesian neural networks, random forests (RFs), and support vector machines (SVMs). These ML algorithms are adept at predicting endpoints characterized by complex and nonlinear relationships. The adoption of deep neural network (DNN) methodologies for silico modeling of ADMET endpoints surged following the Kaggle 'Merck Molecular Activity Challenge' conducted in 2012. This competition aimed to assess the effectiveness of ML methods in predicting 18 distinct ADMET endpoints utilizing datasets of varying sizes (2000–50,000 molecules) sourced from Merck's internal data. The winning submission employed an ensemble approach incorporating DNN, gradient-boosting machine (GBM), and Gaussian process (GP) regression techniques. Merck researchers subsequently published a follow-up study comparing the performance of DNN models against RF models, demonstrating that DNN models frequently outperformed RF in numerous instances. Similarly, in the Tox21 data challenge conducted by the NIH to evaluate computational models for toxicity prediction, DL models excelled and outperformed traditional shallow-learning ML methods.

A distinctive feature of DNNs is their capacity to concurrently train networks that amalgamate different endpoints within a singular model. Through a learning strategy known as inductive transfer learning, multitask DNNs are trained on datasets corresponding to various ADMET endpoints, consolidating them into a unified model. The premise behind multitask DNNs is to facilitate expedited learning and enhance model accuracy by sharing internal representations. Most multitask DNN models utilized for modeling ADMET endpoints employ a 'hard' parameter sharing approach, meaning the hidden layer is shared across all tasks. A general data-driven process for constructing ADMET models is depicted. In a performance benchmarking study evaluating accuracy against 31 assay datasets, Evan et al. demonstrated that multitask DNNs exhibited superior accuracy in predicting ADMET endpoints compared to single-task DNNs and shallow-learning ML methodologies. The multitask DNNs developed in this study encompassed 137,205 compounds and enabled the simultaneous prediction of the majority of ADMET properties, including solubility, permeability, hepatotoxicity, and CYP

inhibition. Similarly, Xu et al. employed multitask learning to successfully predict metabolic stability and drug-drug interaction properties. The performance of multitask models was comparable to traditional single-task approaches while requiring significantly fewer data samples to train the model effectively [139-143].

Molecular Dynamics Simulations

Molecular dynamics (MD) simulations are a computational technique for modeling the time-dependent behavior of molecular systems at an atomic level. MD simulations have become an invaluable tool in modern drug discovery and development, offering insights into the conformational dynamics, stability, and energetics of biomolecular complexes. By providing atomic-level trajectories of molecular motion, MD simulations can elucidate the mechanisms underlying protein-ligand interactions, allosteric effects, and conformational transitions critical for drug binding. MD simulations facilitate the exploration of conformational landscapes that are often inaccessible through traditional experimental methods. Through these simulations, researchers can capture transient interactions and dynamic conformational changes, providing a more comprehensive understanding of the binding modes and affinities of small molecules. Consequently, MD simulations can complement and enhance virtual screening efforts by refining the selection of candidate compounds based on dynamic interactions rather than static docking scores. The integration of MD simulations with advanced sampling techniques has further enhanced their applicability in drug discovery. Techniques such as enhanced sampling methods, including replica exchange MD (REMD), accelerated MD (aMD), and metadynamics, allow researchers to overcome energy barriers and access conformational states that are otherwise infrequently sampled. These methods enable more efficient exploration of complex energy landscapes, yielding a more accurate representation of ligand binding affinities and target interactions.

A notable example of the impact of MD simulations on drug discovery is their application in the development of HIV protease inhibitors. By utilizing MD simulations, researchers have gained insights into the binding dynamics of various inhibitors, enabling the design of compounds with improved binding affinity and selectivity. Additionally, MD simulations have been instrumental in optimizing lead compounds through iterative cycles of design, synthesis, and evaluation, streamlining the drug development process. Moreover, the emergence of high-performance computing resources and cloud-based platforms has accelerated the accessibility and efficiency of MD simulations. This advancement has allowed researchers to perform large-scale simulations, thereby increasing the scope and resolution of investigations into biomolecular interactions. In summary, molecular dynamics simulations serve as a powerful tool in drug discovery, providing detailed insights into the dynamic behavior of molecular systems. Their integration with advanced sampling techniques and computational resources enhances their utility, enabling the exploration of complex molecular landscapes critical for the design of novel therapeutic agents. As the field continues to evolve, the role of MD simulations in informing drug discovery efforts is expected to expand, ultimately contributing to the development of more effective and targeted therapeutics.

Challenges:

Even though artificial intelligence (AI) and machine learning algorithms have advanced significantly in the pharmaceutical industry, there are still several barriers to their successful application and integration into the drug development process and the larger pharmaceutical sector. The poor integration of many data sources is one prominent problem. The variability of datasets—which can include raw data, processed data, metadata, or candidate data—is the source of this difficulty. These datasets need to be gathered and combined for efficient analysis, but there isn't a clear method for doing so at the moment. Because incorrectly formatted data might result in faulty outputs from machine learning algorithms, this kind of integration is essential before starting the drug discovery process. Therefore, more effective methods to compile existing data into data banks prior to starting drug research are urgently needed [144]. The immobility of occupational expertise and skills is another difficulty. Many of the experts working in the pharmaceutical sector today are not qualified or have the necessary abilities to effectively operate AI systems. Few people have the dual expertise that allows AI to be used in a pharmaceutical setting, even though some people are skilled in data science and others focus on molecular chemistry or biology. Developing appropriate algorithms requires an understanding of the underlying chemistry, and successfully utilizing AI applications requires a comprehension of algorithmic concepts [144]. A related challenge is the pharmaceutical industry's mistrust of AI and machine learning, which is mostly caused by a general mistrust of the outcomes and a lack of knowledge of algorithmic approaches, sometimes known as the "black box" syndrome. People who are skeptical might be reluctant to use data produced by AI and machine learning, which would waste time and money and ultimately reduce industry productivity. The pharmaceutical industry's lack of investment in the development of AI technologies is further exacerbated by this mistrust of the technology. Potential investors may be discouraged from investing in AI and machine learning due to doubts about their effectiveness and results in drug development. This hesitation may cause research and development procedures to proceed more slowly and inefficiently than they could, which would impede the advancement of AI in the pharmaceutical industry. For AI to be successfully incorporated into drug development procedures, a number of obstacles that stand in the way of significant development must be removed.

Conclusion

Artificial intelligence (AI) has emerged as a pivotal force in revolutionizing the drug discovery landscape, offering novel approaches that significantly enhance the efficiency, accuracy, and speed of identifying new therapeutic agents. The traditional drug discovery process, characterized by its lengthy timelines and high costs, has long been a bottleneck in the pharmaceutical industry. AI and machine learning (ML) provide innovative solutions by facilitating the analysis of vast datasets and enabling predictive modeling, which allows for more informed decision-making at every stage of drug development. This review highlights the diverse applications of AI in drug discovery, including its role in target identification, lead optimization, and preclinical evaluations. The integration of advanced AI tools, such as deep learning algorithms, natural language

processing, and reinforcement learning, has demonstrated remarkable success in processing complex biological data, predicting drug interactions, and optimizing molecular designs. These technologies have not only accelerated the pace of drug discovery but also improved the likelihood of identifying viable drug candidates that may have otherwise been overlooked. Despite these advancements, several challenges persist. Data integration remains a critical issue, as the heterogeneity of datasets can complicate effective analysis. Moreover, the current workforce in the pharmaceutical industry often lacks the interdisciplinary skills necessary to harness the full potential of AI technologies. Addressing this skills gap is essential for fostering innovation and facilitating collaboration between data scientists and domain experts. Additionally, skepticism regarding the reliability of AI-generated results can hinder acceptance and investment in these technologies, potentially stalling progress. To overcome these challenges, stakeholders in the pharmaceutical industry must prioritize the establishment of standardized data integration methods and invest in training programs that equip professionals with the necessary skills to work with AI systems. By fostering a culture of collaboration and trust in AI technologies, the industry can unlock the transformative potential of AI, paving the way for more efficient and successful drug discovery processes. Moving forward, continued research and development in AI applications will be crucial for enhancing the pharmaceutical landscape and addressing unmet medical needs. The future of drug discovery is undoubtedly intertwined with the advancements in AI, offering promising opportunities for innovation and improved patient outcomes.

References

1. Mullard, A. (2014). New drugs cost us $2.6 billion to develop. Nature Reviews Drug Discovery, 13(12), 877. https://doi.org/10.1038/nrd.2014.203
2. Dowden, H., & Munro, J. (2019). Trends in clinical success rates and therapeutic focus. Nature Reviews Drug Discovery, 18(7), 495–497. https://doi.org/10.1038/d41573-019-00027-7
3. Schneider, G. (2018). Automating drug discovery. Nature Reviews Drug Discovery, 17(2), 97. https://doi.org/10.1038/nrd.2017.202
4. Chen, H., Engkvist, O., Wang, Y., et al. (2018). The rise of deep learning in drug discovery. Drug Discovery Today, 23(6), 1241–1250. https://doi.org/10.1016/j.drudis.2018.01.039
5. Mater, A. C., & Coote, M. L. (2019). Deep learning in chemistry. Journal of Chemical Information and Modeling, 59(6), 2545–2559. https://doi.org/10.1021/acs.jcim.9b00265
6. Vamathevan, J., Clark, D., Czodrowski, P., et al. (2019). Applications of machine learning in drug discovery and development. Nature Reviews Drug Discovery, 18(6), 463–477. https://doi.org/10.1038/s41573-019-0024-5
7. Paul, D., Sanap, G., Shenoy, S., et al. (2020). Artificial intelligence in drug discovery and development. Drug Discovery Today, 26(1), 80–93. https://doi.org/10.1016/j.drudis.2020.07.020
8. Stumpfe, D., & Bajorath, J. (2020). Current trends, overlooked issues, and unmet challenges in virtual screening. Journal of Chemical Information and Modeling, 60(9), 4112–4115. https://doi.org/10.1021/acs.jcim.0c00535

9. Schneider, P., Walters, W. P., Plowright, A. T., et al. (2020). Rethinking drug design in the artificial intelligence era. Nature Reviews Drug Discovery, 19(5), 353–364. https://doi.org/10.1038/s41573-020-00083-5
10. Boström, J., Brown, D. G., Young, R. J., et al. (2018). Expanding the medicinal chemistry synthetic toolbox. Nature Reviews Drug Discovery, 17(10), 709–727. https://doi.org/10.1038/nrd.2018.141
11. Strokach, A., Becerra, D., Corbi-Verge, C., et al. (2020). Fast and flexible protein design using deep graph neural networks. Cell Systems, 11(4), 402–411. https://doi.org/10.1016/j.cels.2020.09.003
12. Rantanen, J., & Khinast, J. (2015). The future of pharmaceutical manufacturing sciences. Journal of Pharmaceutical Sciences, 104(11), 3612–3638. https://doi.org/10.1002/jps.24594
13. Greenhill, A. T., & Edmunds, B. R. (2020). A primer of artificial intelligence in medicine. Technology Innovations in Gastrointestinal Endoscopy, 22(2), 85–89. https://doi.org/10.1016/j.tgie.2019.150642
14. Solanki, P., Baldaniya, D., Jogani, D., et al. (2021). Artificial intelligence: New age of transformation in petroleum upstream. Petroleum Research. https://doi.org/10.1016/j.ptlrs.2021.07.002
15. Kshirsagar, A., & Shah, M. (2021). Anatomized study of security solutions for multimedia: Deep learning-enabled authentication, cryptography and information hiding. In Advances in Security Solutions for Multimedia (pp. 1–19). https://doi.org/10.1088/978-0-7503-3735-9CH7
16. Anthony, C. C. (2016). Big data in medicine: The upcoming artificial intelligence. Progress in Pediatric Cardiology, 43, 91–94. https://doi.org/10.1016/j.ppedcard.2016.08.021
17. Hughes, J. P., Rees, S., Kalindjian, S. B., et al. (2011). Principles of early drug discovery. British Journal of Pharmacology, 162(6), 1239–1249.
18. Pereira, D. A., & Williams, J. A. (2007). Origin and evolution of high throughput screening. British Journal of Pharmacology, 152(1), 53–61.
19. Bender, A., Bojanic, D., Davies, J. W., et al. (2008). Which aspects of HTS are empirically correlated with downstream success? Current Opinion in Drug Discovery & Development, 11(3), 327.
20. Wang, Y., Bryant, S. H., Cheng, T., et al. (2017). PubChem bioassay: 2017 update. Nucleic Acids Research, 45(D1), D955–D963.
21. Sterling, T., & Irwin, J. J. (2015). ZINC 15—Ligand discovery for everyone. Journal of Chemical Information and Modeling, 55(11), 2324–2337.
22. Kim, S. (2016). Getting the most out of PubChem for virtual screening. Expert Opinion on Drug Discovery, 11(9), 843–855.
23. Scior, T., Bender, A., Tresadern, G., et al. (2012). Recognizing pitfalls in virtual screening: A critical review. Journal of Chemical Information and Modeling, 52(4), 867–881.
24. Salahudeen, M. S., & Nishtala, P. S. (2017). An overview of pharmacodynamic modeling, ligand-binding approach, and its application in clinical practice. Saudi Pharmaceutical Journal, 25(2), 165–175.
25. Hu, Y., & Bajorath, J. (2013). Compound promiscuity: What can we learn from current data? Drug Discovery Today, 18(13–14), 644–650.
26. Yusof, I., Shah, F., Hashimoto, T., et al. (2014). Finding the rules for successful drug optimization. Drug Discovery Today, 19(5), 680–687.
27. Nicolaou, C. A., & Brown, N. (2013). Multi-objective optimization methods in drug design. Drug Discovery Today: Technologies, 10(3), e427–e435.

28. Muratov, E. N., Bajorath, J., Sheridan, R. P., et al. (2020). QSAR without borders. Chemical Society Reviews, 49(11), 3525–3564.
29. Schneider, G., & Fechner, U. (2005). Computer-based de novo design of drug-like molecules. Nature Reviews Drug Discovery, 4(8), 649–663.
30. Dobson, C. M. (2004). Chemical space and biology. Nature, 432(7019), 824–828.
31. Sliwoski, G., Kothiwale, S., Meiler, J., et al. (2014). Computational methods in drug discovery. Pharmacological Reviews, 66(1), 334–395.
32. Van Drie, J. H. (2007). Computer-aided drug design: The next 20 years. Journal of Computer-Aided Molecular Design, 21(10), 591–601.
33. Jiménez-Luna, J., Grisoni, F., & Schneider, G. (2020). Drug discovery with explainable artificial intelligence. Nature Machine Intelligence, 2(10), 573–584.
34. Göller, A. H., Kuhnke, L., Montanari, F., et al. (2020). Bayer's in silico ADMET platform: A journey of machine learning over the past two decades. Drug Discovery Today, 25(9), 1702–1709. https://doi.org/10.1016/j.drudis.2020.07.001
35. Abu-Elezz, I., Hassan, A., Nazeemudeen, A., et al. (2020). The benefits and threats of blockchain technology in healthcare: A scoping review. International Journal of Medical Informatics, 142, Article 104246. https://doi.org/10.1016/j.ijmedinf.2020.104246
36. Park, B. J., Choi, H. J., Moon, S. J., et al. (2018). Pharmaceutical applications of 3D printing technology: Current understanding and future perspectives. Journal of Pharmaceutical Investigation, 49(6), 575–585. https://doi.org/10.1007/S40005-018-00414-Y
37. Zimmerling, A., & Chen, X. (2020). Bioprinting for combating infectious diseases. Bioprinting, 20, e00104. https://doi.org/10.1016/j.bprint.2020.e00104
38. Ma, J., Sheridan, R. P., Liaw, A., et al. (2015). Deep neural nets as a method for quantitative structure–activity relationships. Journal of Chemical Information and Modeling, 55(2), 263–274. https://doi.org/10.1021/ci5006002
39. Lavecchia, A. (2015). Machine-learning approaches in drug discovery: Methods and applications. Drug Discovery Today, 20(3), 318–331. https://doi.org/10.1016/j.drudis.2014.11.007
40. Krizhevsky, A., Sutskever, I., & Hinton, G. E. (2012). Imagenet classification with deep convolutional neural networks. In Advances in Neural Information Processing Systems (Vol. 25, pp. 1097–1105).
41. Alom, M. Z., Taha, T. M., Yakopcic, C., et al. (2018). The history began from AlexNet: A comprehensive survey on deep learning approaches. arXiv preprint arXiv:1803.01164.
42. Öztürk, H., Özgür, A., Schwaller, P., et al. (2020). Exploring chemical space using natural language processing methodologies for drug discovery. Drug Discovery Today, 25(4), 689–705. https://doi.org/10.1016/j.drudis.2019.12.010
43. Jiménez-Luna, J., Grisoni, F., Weskamp, N., et al. (2021). Artificial intelligence in drug discovery: Recent advances and future perspectives. Expert Opinion on Drug Discovery, 1–11. https://doi.org/10.1080/17460441.2021.1872984
44. Zhavoronkov, A., Ivanenkov, Y. A., Aliper, A., et al. (2019). Deep learning enables rapid identification of potent DDR1 kinase inhibitors. Nature

Biotechnology, 37(9), 1038–1040. https://doi.org/10.1038/s41587-019-0238-5
45. Stokes, J. M., Yang, K., Swanson, K., et al. (2020). A deep learning approach to antibiotic discovery. Cell, 180(4), 688–702. https://doi.org/10.1016/j.cell.2020.01.007
46. Chuang, K. V., Gunsalus, L. M., & Keiser, M. J. (2020). Learning molecular representations for medicinal chemistry: Miniperspective. Journal of Medicinal Chemistry, 63(16), 8705–8722. https://doi.org/10.1021/acs.jmedchem.0c00601
47. Mayr, A., Klambauer, G., Unterthiner, T., et al. (2016). Deeptox: Toxicity prediction using deep learning. Frontiers in Environmental Science, 3, 80. https://doi.org/10.3389/fenvs.2016.00080
48. Andrade, R. J., Chalasani, N., Björnsson, E. S., et al. (2019). Drug-induced liver injury. Nature Reviews Disease Primers, 5(1), 1–22. https://doi.org/10.1038/s41572-019-0070-8
49. Elton, D. C., Boukouvalas, Z., Fuge, M. D., et al. (2019). Deep learning for molecular design: A review of the state of the art. Molecular Systems Design & Engineering, 4(4), 828–849. https://doi.org/10.1039/C9ME00011F
50. Mercado, R., Rastemo, T., Lindelöf, E., et al. (2020). Practical notes on building molecular graph generative models. Applied AI Letters, 1(2). https://doi.org/10.1002/ail2.15
51. Schaduangrat, N., Lampa, S., Simeon, S., et al. (2020). Towards reproducible computational drug discovery. Journal of Chemical Information, 12(1), 9. https://doi.org/10.3390/jcce12010009
52. Bender, A., & Cortes-Ciriano, I. (2021). Artificial intelligence in drug discovery: What is realistic, what are illusions? Part 1: Ways to make an impact, and why we are not there yet. Drug Discovery Today, 26(2), 511–524. https://doi.org/10.1016/j.drudis.2020.11.021
53. Bender, A., & Cortes-Ciriano, I. (2021). Artificial intelligence in drug discovery: What is realistic, what are illusions? Part 2: A discussion of chemical and biological data used for AI in drug discovery. Drug Discovery Today, 26(4), 1040–1052. https://doi.org/10.1016/j.drudis.2020.11.022
54. Walters, W. P., & Barzilay, R. (2021). Critical assessment of AI in drug discovery. Expert Opinion on Drug Discovery, 1–11. https://doi.org/10.1080/17460441.2021.1872984
55. Elbadawi M, Gaisford S, Basit AW. Advanced machine-learning techniques in drug discovery. Drug Discov Today 2021;26(3):769-777. doi:10.1016/j.drudis.2020.12.003.
56. Bender A, Cortes-Ciriano I. Artificial intelligence in drug discovery: what is realistic, what are illusions? Part 2: a discussion of chemical and biological data. Drug Discov Today 2021;26(4):1040-1052. doi:10.1016/j.drudis.2020.11.037.
57. Reker D. Practical considerations for active machine learning in drug discovery. Drug Discov Today Technol 2019;32-33:73-79. doi:10.1016/j.ddtec.2020.06.001.
58. Margulis E, Dagan-Wiener A, Ives RS, et al. Intense bitterness of molecules: machine learning for expediting drug discovery. Comput Struct Biotechnol J 2020;19:568-576. doi:10.1016/j.csbj.2020.12.030.

59. Raschka S, Kaufman B. Machine learning and AI-based approaches for bioactive ligand discovery and GPCR-ligand recognition. Methods 2020;180:89-110. doi:10.1016/j.ymeth.2020.06.016.
60. Raschka S. Automated discovery of GPCR bioactive ligands. Curr Opin Struct Biol 2019;55:17-24. doi:10.1016/j.sbi.2019.02.011.
61. Rantanen J, Khinast J. The future of pharmaceutical manufacturing sciences. J Pharm Sci 2015;104(11):3612-3638. doi:10.1002/jps.24594.
62. Turki T, Taguchi YH. Machine learning algorithms for predicting drugs–tissues relationships. Expert Syst Appl 2019;127:167-186. doi:10.1016/j.eswa.2019.02.013.
63. Popova M, Isayev O, Tropsha A. Deep reinforcement learning for de novo drug design. Sci Adv 2018;4(7) . doi:10.1126/sciadv.aap7885.
64. Taroni JN, Grayson PC, Hu Q, et al. MultiPLIER: a transfer learning framework for transcriptomics reveals systemic features of rare disease. Cell Syst 2019;8(5):380-394.e4. doi:10.1016/j.cels.2019.04.003.
65. Li L, He X, Borgwardt K. Multi-target drug repositioning by bipartite block-wise sparse multi-task learning. BMC Syst Biol 2018;12(Suppl 4):55. doi:10.1186/s12918-018-0569-7.
66. Weng Y, Lin C, Zeng X, et al. Drug Target interaction prediction using multi-task learning and co-attention. Proceedings of the IEEE International Conference on Bioinformatics and Biomedicine (BIBM), 2019;528-533. doi:10.1109/BIBM47256.2019.8983254. n
67. Han L, Zhang Y. Learning multi-level task groups in multi-task learning. Proceedings of the Twenty-Ninth AAAI Conference on Artificial Intelligence, 2015. Available from: https://www.aaai.org/ocs/index.php/AAAI/AAAI15/paper/view/9510.
68. Mak, K. K., & Pichika, M. R. (2019). Artificial intelligence in drug development: Present status and future prospects. Drug Discovery Today, 24(3), 773-780. https://doi.org/10.1016/j.drudis.2018.11.014
69. Paul, D., Sanap, G., Shenoy, S., et al. (2021). Artificial intelligence in drug discovery and development. Drug Discovery Today, 26(1), 80-93. https://doi.org/10.1016/j.drudis.2020.10.010
70. Dnyaneshwar, K., Gaurav, S., Debleena, P., et al. (2020). Artificial intelligence in the pharmaceutical sector: Current scene and future prospect. In The future of pharmaceutical product development and research (pp. 73-107). Elsevier. https://doi.org/10.1016/B978-0-12-814455-8.00003-7
71. Cui, R. B., & Zhu, F. (2021). Ultrasound modified polysaccharides: A review of structure, physicochemical properties, biological activities and food applications. Trends in Food Science & Technology, 107, 491-508. https://doi.org/10.1016/J.TIFS.2020.11.018
72. Lusci, A., Pollastri, G., & Baldi, P. (2013). Deep architectures and deep learning in chemoinformatics: The prediction of aqueous solubility for drug-like molecules. Journal of Chemical Information and Modeling, 53(7), 1563-1575. https://doi.org/10.1021/ci400187y
73. Polykovskiy, D., Zhebrak, A., Vetrov, D., et al. (2018). Entangled conditional adversarial autoencoder for de novo drug discovery. Molecular Pharmaceutics, 15(10), 4398-4405. https://doi.org/10.1021/acs.molpharmaceut.8b00839
74. Daynac, M., Cortes-Cabrera, A., & Prieto, J. M. (2015). Application of artificial intelligence to the prediction of the antimicrobial activity of essential oils.

Evidence-Based Complementary and Alternative Medicine, 2015, Article 561024. https://doi.org/10.1155/2015/561024
75. Pu, L., Naderi, M., Liu, T., et al. (2019). eToxPred: A machine learning-based approach to estimate the toxicity of drug candidates. BMC Pharmacology and Toxicology, 20(1), 2. https://doi.org/10.1186/s40360-018-0282-6
76. Kadurin, I., Rothwell, S., Ferron, L., et al. (2017). Investigation of the proteolytic cleavage of α2δ subunits: A mechanistic switch from inhibition to activation of voltage-gated calcium channels? Biophysical Journal, 112(3), 244a. https://doi.org/10.1016/j.bpj.2016.11.1335
77. Maram, Y., & Hamdy, H. (2021). The role of artificial intelligence in revealing the results of the interaction of biological materials with each other or with chemicals. Materials Today: Proceedings, 45, 4954-4959. https://doi.org/10.1016/j.matpr.2021.01.387
78. Unterthiner, T., Mayr, A., Klambauer, G., et al. (2015). Toxicity prediction using deep learning. Retrieved from https://arxiv.org/abs/1503.01445v1
79. Avdagic, Z., Begic Fazlic, L., & Konjicija, S. (2009). Optimized detection of tar content in the manufacturing process using adaptive neuro-fuzzy inference systems. Studies in Health Technology and Informatics, 150, 615-619.
80. Zhavoronkov, A., Ivanenkov, Y. A., Aliper, A., et al. (2019). Deep learning enables rapid identification of potent DDR1 kinase inhibitors. Nature Biotechnology, 37(9), 1038-1040. https://doi.org/10.1038/s41587-019-0224-x
81. Lee, H., & Kim, W. (2019). Comparison of target features for predicting drug-target interactions by deep neural network based on large-scale drug-induced transcriptome data. Pharmaceutics, 11(8), 377. https://doi.org/10.3390/pharmaceutics11080377
82. Putin, E., Asadulaev, A., Ivanenkov, Y., et al. (2018). Reinforced adversarial neural computer for de novo molecular design. Journal of Chemical Information and Modeling, 58(6), 1194-1204. https://doi.org/10.1021/acs.jcim.7b00690
83. Sellwood, M. A., Ahmed, M., Segler, M. H., & Brown, N. (2018). Artificial intelligence in drug discovery. *Future Medicinal Chemistry, 10*(20), 2025-2028. https://doi.org/10.4155/fmc-2018-0168
84. Paul, D., Sanap, G., Shenoy, S., Kalyane, D., Kalia, K., & Tekade, R. K. (2021). Artificial intelligence in drug discovery and development. *Drug Discovery Today, 26*(1), 80-93. https://doi.org/10.1016/j.drudis.2020.10.010
85. Edfeldt, F. N., Folmer, R. H., & Breeze, A. L. (2011). Fragment screening to predict druggability (ligandability) and lead discovery success. *Drug Discovery Today, 16*(6), 284-287. https://doi.org/10.1016/j.drudis.2011.02.003
86. Vukovic, S., & Huggins, D. J. (2018). Quantitative metrics for drug-target ligandability. *Drug Discovery Today, 23*(7), 1258-1266. https://doi.org/10.1016/j.drudis.2018.01.003
87. UniProt. (2021). UniProt: The universal protein knowledgebase in 2021. *Nucleic Acids Research, 49*(D1), D480-D489. https://doi.org/10.1093/nar/gkaa1099
88. Berman, H. M., Westbrook, J., Feng, Z., Gilliland, G., Bhat, T. N., Weissig, H., & others. (2000). The Protein Data Bank. *Nucleic Acids Research, 28*(1), 235-242. https://doi.org/10.1093/nar/28.1.235

89. Hillisch, A., Pineda, L. F., & Hilgenfeld, R. (2004). Utility of homology models in the drug discovery process. *Drug Discovery Today, 9*(14), 659-669. https://doi.org/10.1016/j.drudis.2004.06.002
90. Kuhlman, B., & Bradley, P. (2019). Advances in protein structure prediction and design. *Nature Reviews Molecular Cell Biology, 20*(11), 681-697. https://doi.org/10.1038/s41580-019-0150-5
91. Kinch, L. N., Kryshtafovych, A., Monastyrskyy, B., & Grishin, N. V. (2019). CASP 13 target classification into tertiary structure prediction categories. *Proteins, 87*(10), 1021-1036. https://doi.org/10.1002/prot.25725
92. Jumper, J., Evans, R., Pritzel, A., Green, T., Figurnov, M., Tunyasuvunakool, K., & others. (2020). High accuracy protein structure prediction using deep learning. In *Fourteenth Critical Assessment of Techniques for Protein Structure Prediction* (pp. 22–24). Davis: Protein Structure Prediction Center.
93. The AlphaFold Team. (2021). AlphaFold: A solution to a 50-year-old grand challenge in biology. https://deepmind.com/blog/article/alphafold-a-solution-to-a-50-year-old-grand-challenge-in-biology
94. Jumper, J., Evans, R., Pritzel, A., Green, T., Figurnov, M., Ronneberger, O., & others. (2021). Highly accurate protein structure prediction with AlphaFold. *Nature, 596*(7873), 583-589. https://doi.org/10.1038/s41586-020-2649-2
95. Tunyasuvunakool, K., Adler, J., Wu, Z., Green, T., Zielinski, M., Zidek, A., & others. (2021). Highly accurate protein structure prediction for the human proteome. *Nature, 596*(7873), 590-596. https://doi.org/10.1038/s41586-020-2649-2
96. Baek, M., DiMaio, F., Anishchenko, I., Dauparas, J., Ovchinnikov, S., Lee, G. R., & others. (2021). Accurate prediction of protein structures and interactions using a three-track neural network. *Science, 373*(6559), 871-876. https://doi.org/10.1126/science.abj8754
97. Andreeva, A., Kulesha, E., Gough, J., & Murzin, A. G. (2020). The SCOP database in 2020: Expanded classification of representative family and superfamily domains of known protein structures. *Nucleic Acids Research, 48*(D1), D376-D382. https://doi.org/10.1093/nar/gkz900
98. Grant, A., Lee, D., & Orengo, C. (2004). Progress towards mapping the universe of protein folds. *Genome Biology, 5*(9), 107. https://doi.org/10.1186/gb-2004-5-9-107
99. Ehrt, C., Brinkjost, T., & Koch, O. (2016). Impact of binding site comparisons on medicinal chemistry and rational molecular design. *Journal of Medicinal Chemistry, 59*(9), 4121-4151. https://doi.org/10.1021/acs.jmedchem.5b01283
100. Simonovsky, M., & Meyers, J. (2020). DeeplyTough: Learning structural comparison of protein binding sites. *Journal of Chemical Information and Modeling, 60*(5), 2356-2366. https://doi.org/10.1021/acs.jcim.0c00111
101. Zhu, T., Cao, S., Su, P. C., Patel, R., Shah, D., Chokshi, H. B., & others. (2013). Hit identification and optimization in virtual screening: Practical recommendations based on a critical literature analysis. *Journal of Medicinal Chemistry, 56*(16), 6560-6572. https://doi.org/10.1021/jm401358t
102. Ripphausen, P., Nisius, B., Peltason, L., & Bajorath, J. (2010). Quo vadis, virtual screening? A comprehensive survey of prospective applications. Journal of Medicinal Chemistry, 53(23), 8461-8467. https://doi.org/10.1021/jm100646j

103. Damm-Ganamet, K. L., Arora, N., Becart, S., Edwards, J. P., Lebsack, A. D., McAllister, H. M., et al. (2019). Accelerating lead identification by high throughput virtual screening: Prospective case studies from the pharmaceutical industry. Journal of Chemical Information and Modeling, 59(5), 2046-2062. https://doi.org/10.1021/acs.jcim.9b00051
104. Gorgulla, C., Boeszoermenyi, A., Wang, Z. F., Fischer, P. D., Coote, P. W., Padmanabha Das, K. M., et al. (2020). An open-source drug discovery platform enables ultra-large virtual screens. Nature, 580(7805), 663-668. https://doi.org/10.1038/s41586-020-2155-5
105. Wang, Y., Bryant, S. H., Cheng, T., Wang, J., Gindulyte, A., Shoemaker, B. A., et al. (2017). PubChem bioassay: 2017 update. Nucleic Acids Research, 45(D1), D955-D963. https://doi.org/10.1093/nar/gkw1019
106. Mendez, D., Gaulton, A., Bento, A. P., Chambers, J., De Veij, M., Felix, E., et al. (2019). ChEMBL: Towards direct deposition of bioassay data. Nucleic Acids Research, 47(D1), D930-D940. https://doi.org/10.1093/nar/gky1078
107. Zhang, L., Fourches, D., Sedykh, A., Zhu, H., Golbraikh, A., Ekins, S., et al. (2013). Discovery of novel antimalarial compounds enabled by QSAR-based virtual screening. Journal of Chemical Information and Modeling, 53(3), 475-492. https://doi.org/10.1021/ci300677s
108. Anantpadma, M., Lane, T., Zorn, K. M., Lingerfelt, M. A., Clark, A. M., Freundlich, J. S., et al. (2019). Ebola virus Bayesian machine learning models enable new in vitro leads. ACS Omega, 4(4), 2353-2361. https://doi.org/10.1021/acsomega.8b03395
109. Chen, X., Xie, W., Yang, Y., Hua, Y., Xing, G., Liang, L., et al. (2020). Discovery of dual FGFR4 and EGFR inhibitors by machine learning and biological evaluation. Journal of Chemical Information and Modeling, 60(10), 4640-4652. https://doi.org/10.1021/acs.jcim.0c00749
110. Donlin, M. J., Lane, T. R., Riabova, O., Lepioshkin, A., Xu, E., Lin, J., et al. (2021). Discovery of 5-nitro-6-thiocyanatopyrimidines as inhibitors of Cryptococcus neoformans and Cryptococcus gattii. ACS Medicinal Chemistry Letters, 12(7), 774-781. https://doi.org/10.1021/acsmedchemlett.0c00609
111. Liu, Z., Huang, D., Zheng, S., Song, Y., Liu, B., Sun, J., et al. (2021). Deep learning enables discovery of highly potent anti-osteoporosis natural products. European Journal of Medicinal Chemistry, 210, Article 112982. https://doi.org/10.1016/j.ejmech.2020.112982
112. Verkhivker, G. M., Bouzida, D., Gehlhaar, D. K., Rejto, P. A., Arthurs, S., Colson, A. B., et al. (2000). Deciphering common failures in molecular docking of ligand-protein complexes. Journal of Computer-Aided Molecular Design, 14(8), 731-751. https://doi.org/10.1023/A:1008995515685
113. Wicht, K. J., Combrinck, J. M., Smith, P. J., & Egan, T. J. (2015). Bayesian models trained with HTS data for predicting beta-haematin inhibition and in vitro antimalarial activity. Bioorganic & Medicinal Chemistry, 23(18), 5210-5217. https://doi.org/10.1016/j.bmc.2015.07.017
114. Yang, M., Tao, B., Chen, C., Jia, W., Sun, S., Zhang, T., et al. (2019). Machine learning models based on molecular fingerprints and an extreme gradient boosting method lead to the discovery of JAK2 inhibitors. Journal of Chemical Information and Modeling, 59(10), 5002-5012. https://doi.org/10.1021/acs.jcim.9b00408

115. Liu, Z., Du, J., Fang, J., Yin, Y., Xu, G., & Xie, L. (2019). Deepscreening: A deep learning-based screening web server for accelerating drug discovery. Database (Oxford), 2019, baz104. https://doi.org/10.1093/database/baz104
116. Soufan, O., Ba-Alawi, W., Magana-Mora, A., Essack, M., & Bajic, V. B. (2018). DPubChem: A web tool for QSAR modeling and high-throughput virtual screening. Scientific Reports, 8, 9110. https://doi.org/10.1038/s41598-018-27480-3
117. Kitchen, D. B., Decornez, H., Furr, J. R., & Bajorath, J. (2004). Docking and scoring in virtual screening for drug discovery: Methods and applications. Nature Reviews Drug Discovery, 3(11), 935-949. https://doi.org/10.1038/nrd1540
118. Halperin, I., Ma, B., Wolfson, H., & Nussinov, R. (2002). Principles of docking: An overview of search algorithms and a guide to scoring functions. Proteins, 47(4), 409-443. https://doi.org/10.1002/prot.10154
119. Pereira, J. C., Caffarena, E. R., & Dos Santos, C. N. (2016). Boosting docking-based virtual screening with deep learning. Journal of Chemical Information and Modeling, 56(12), 2495-2506. https://doi.org/10.1021/acs.jcim.6b00455
120. Wang, Z., Sun, H., Yao, X., Li, D., Xu, L., Li, Y., et al. (2016). Comprehensive evaluation of ten docking programs on a diverse set of protein-ligand complexes: The prediction accuracy of sampling power and scoring power. Physical Chemistry Chemical Physics, 18(18), 12964-12975. https://doi.org/10.1039/C6CP01312A
121. Wallach, I., Dzamba, M., & Heifets, A. (2015). AtomNet: A deep, convolutional neural network for bioactivity prediction in structure-based drug discovery. arXiv, arXiv:1510.02855v1. https://arxiv.org/abs/1510.02855
122. Gilson, M. K., Given, J. A., Bush, B. L., & McCammon, J. A. (1997). The statistical-thermodynamic basis for computation of binding affinities: A critical review. Biophysical Journal, 72(3), 1047–1069.
123. Ashtawy, H. M., & Mahapatra, N. R. (2015). Machine-learning scoring functions for identifying native poses of ligands docked to known and novel proteins. BMC Bioinformatics, 16(Suppl. 6), S3.
124. Bao, J., He, X., & Zhang, J. Z. H. (2021). DeepBSP—a machine learning method for accurate prediction of protein-ligand docking structures. Journal of Chemical Information and Modeling, 61(5), 2231–2240.
125. 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.
126. Adeshina, Y. O., Deeds, E. J., & Karanicolas, J. (2020). Machine learning classification can reduce false positives in structure-based virtual screening. Proceedings of the National Academy of Sciences of the United States of America, 117(36), 18477–18488.
127. Ballester, P. J., & Mitchell, J. B. (2010). A machine learning approach to predicting protein-ligand binding affinity with applications to molecular docking. Bioinformatics, 26(10), 1169–1175.
128. Stepniewska-Dziubinska, M. M., Zielenkiewicz, P., & Siedlecki, P. (2018). Development and evaluation of a deep learning model for protein-ligand binding affinity prediction. Bioinformatics, 34(20), 3666–3674.

129. Zheng, L., Fan, J., & Mu, Y. (2019). Onionnet: A multiple-layer intermolecular-contact-based convolutional neural network for protein-ligand binding affinity prediction. ACS Omega, 4(27), 15956–15965.
130. Durrant, J. D., & McCammon, J. A. (2011). NNScore 2.0: A neural-network receptor-ligand scoring function. Journal of Chemical Information and Modeling, 51(12), 2897–2903.
131. Wang, C., & Zhang, Y. (2017). Improving scoring-docking-screening powers of protein–ligand scoring functions using random forest. Journal of Computational Chemistry, 38(3), 169–177.
132. Shen, C., Hu, Y., Wang, Z., Zhang, X. J., Zhong, H. Y., Wang, G. A., et al. (2021). Can machine learning consistently improve the scoring power of classical scoring functions? Insights into the role of machine learning in scoring functions. Briefings in Bioinformatics, 22(2), 497–514.
133. Wang, R., Fang, X., Lu, Y., Yang, C. Y., & Wang, S. (2005). The PDBbind database: Methodologies and updates. Journal of Medicinal Chemistry, 48(12), 4111–4119.
134. Benson, M. L., Smith, R. D., Khazanov, N. A., Dimcheff, B., Beaver, J., Dresslar, P., et al. (2008). Binding MOAD, a high-quality protein-ligand database. Nucleic Acids Research, 36(Database issue), D674–D678.
135. Li, H. J., Sze, K. H., Lu, G., & Ballester, P. J. (2021). Machine-learning scoring functions for structure-based virtual screening. Wiley Interdisciplinary Reviews: Computational Molecular Science, 11(3), Article e1478.
136. Lyu, J., Wang, S., Balius, T. E., Singh, I., Levit, A., Moroz, Y. S., et al. (2019). Ultra-large library docking for discovering new chemotypes. Nature, 566(7745), 224–229.
137. Friesner, R. A., Banks, J. L., Murphy, R. B., Halgren, T. A., Klicic, J. J., Mainz, D. T., et al. (2004). Glide: A new approach for rapid, accurate docking and scoring. 1. Method and assessment of docking accuracy. Journal of Medicinal Chemistry, 47(7), 1739–1749.
138. Ramsundar, B., Eastman, P., Walters, P., Pande, V., Leswing, K., & Wu, Z. (2019). Deep learning for the life sciences. O'Reilly Media.
139. Darvas, F. (1974). Application of the sequential simplex method in designing drug analogs. Journal of Medicinal Chemistry, 17(8), 799–804.
140. Böhm, H. J. (1992). Ludi: Rule-based automatic design of new substituents for enzyme inhibitor leads. Journal of Computer-Aided Molecular Design, 6(6), 593–606.
141. Elbadawi, M., Gaisford, S., & Basit, A. W. (2021). Advanced machine-learning techniques in drug discovery. Drug Discovery Today, 26(4), 769–777.
142. Meyers, J., Fabian, B., & Brown, N. (2021). De novo molecular design and generative models. Drug Discovery Today, 26(11), 2707–2715.
143. Young, R. J., Green, D. V. S., Luscombe, C. N., & Hill, A. P. (2011). Getting physical in drug discovery II: The impact of chromatographic hydrophobicity measurements and aromaticity. *Drug Discovery Today, 16*(18), 822–830.
144. Henstock, P. V. (2019). Artificial intelligence for pharma: Time for internal investment. *Trends in Pharmacological Sciences, 40*(8), 543–546. https://doi.org/10.1016/j.tips.2019.05.003

الذكاء الاصطناعي في اكتشاف الأدوية: التطبيقات الحالية والاتجاهات المستقبلية

الملخص:

الخلفية :عملية اكتشاف الأدوية معقدة وتستغرق وقتًا طويلاً وتكلف الكثير، وتعتمد تقليديًا على أساليب التجريب والخطأ. لقد ظهر دمج الذكاء الاصطناعي (AI) وتعلم الآلة (ML) كحل تحويلي، يعزز الكفاءة والدقة في تحديد مرشحي الأدوية المحتملين.

الهدف :تهدف هذه المراجعة إلى استكشاف التطبيقات الحالية للذكاء الاصطناعي في اكتشاف الأدوية، وتسليط الضوء على الأدوات المستخدمة في العملية، ومناقشة التحديات المرتبطة بذلك.

الطرق :تم إجراء مراجعة شاملة للأدبيات، مع التركيز على المقالات التي تم مراجعتها من قبل الأقران، والدراسات السريرية، وتقارير الحالة التي توضح تطبيق الذكاء الاصطناعي وتعلم الآلة في مختلف مراحل اكتشاف الأدوية، بما في ذلك تحديد الأهداف، وتحسين القادة، والتقييم قبل السريري.

النتائج :تحدد المراجعة عدة تطبيقات للذكاء الاصطناعي، مثل النمذجة التنبؤية، وتصميم الجزيئات، والفحص الافتراضي، والتي تسريع بشكل كبير من عملية الاكتشاف. كانت الأدوات مثل التعلم العميق، ومعالجة اللغة الطبيعية، والتعلم المعزز ضرورية في تحليل مجموعات البيانات الكبيرة والتنبؤ بتفاعلات الأدوية. ومع ذلك، لا تزال هناك تحديات، بما في ذلك مشاكل تكامل البيانات، والفجوات في المهارات بين المحترفين، والشكوك بشأن نتائج الذكاء الاصطناعي.

الخاتمة :لدى الذكاء الاصطناعي القدرة على إحداث ثورة في اكتشاف الأدوية من خلال تبسيط العمليات وتحسين الدقة. ومع ذلك، من الضروري معالجة التحديات المتعلقة بإدارة البيانات، وتدريب القوى العاملة، والثقة في أنظمة الذكاء الاصطناعي لتحقيق إمكاناته الكاملة في صناعة الأدوية.

الكلمات المفتاحية :الذكاء الاصطناعي، اكتشاف الأدوية، تعلم الآلة، النمذجة التنبؤية، صناعة الأدوية.