JETIR.ORG ISSN: 2349-5162 | ESTD Year : 2014 | Monthly Issue
JOURNAL OF EMERGING TECHNOLOGIES AND INNOVATIVE RESEARCH (JETIR)
An International Scholarly Open Access, Peer-reviewed, Refereed Journal

Artificial intelligence to deep learning: machine intelligence approach for drug discovery

Ram Lakhan Rajpoot, Nitin Kumar, Vivek pal

Student, Department of Biotechnology, R.R. Institute of modern technology, (AKTU), Lucknow Uttar Pradesh, India.

Abstract

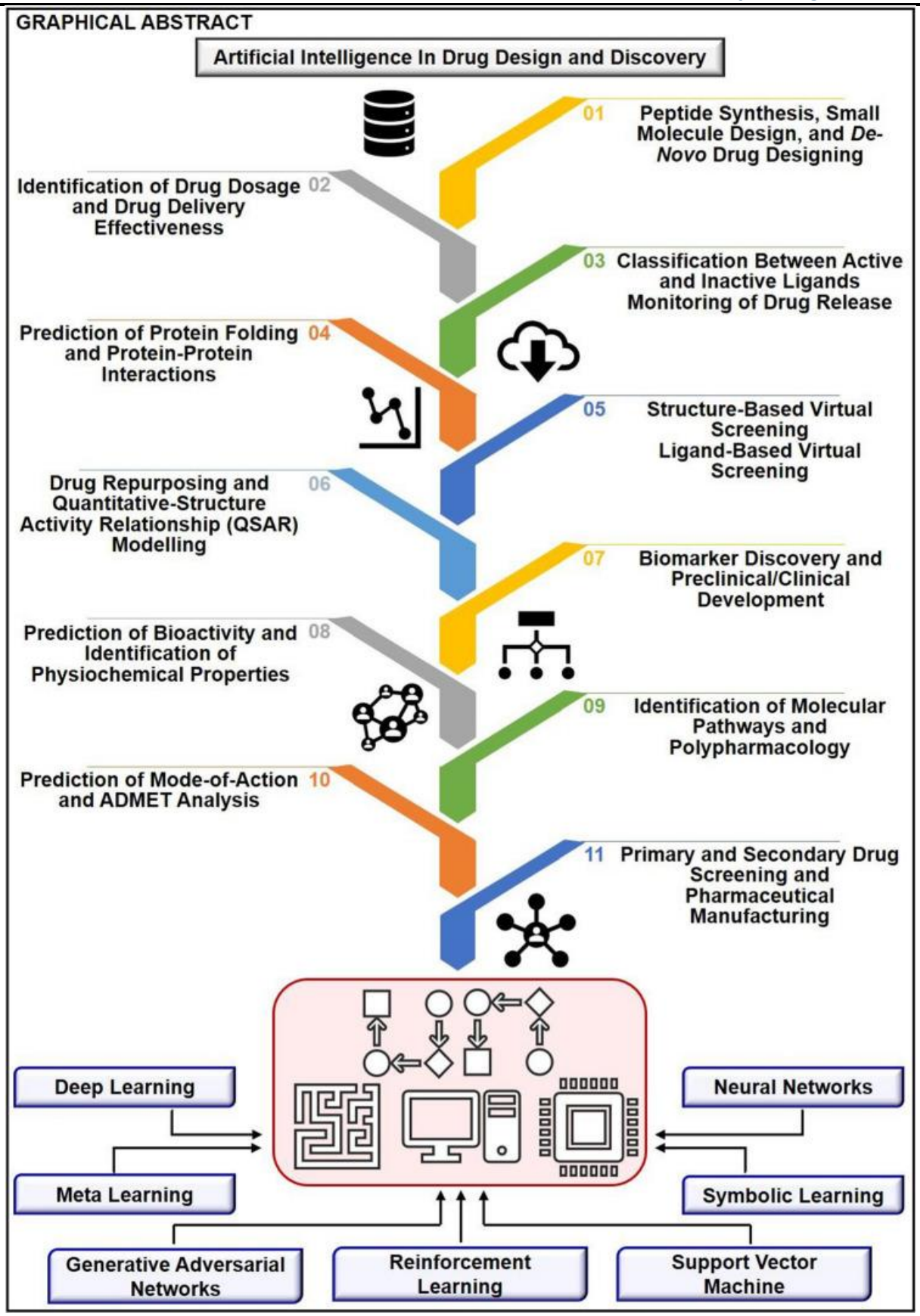
Drug designing and development is an important area of research for pharmaceutical companies and chemical scientists. However, low efficacy, off-target delivery, time consumption, and high cost impose a hurdle and challenges that impact drug design and discovery. Further, complex and big data from genomics, proteomics, microarray data, and clinical trials also impose an obstacle in the drug discovery pipeline. Artificial intelligence and machine learning technology play a crucial role in drug discovery and development. In other words, artificial neural networks and deep learning algorithms have modernized the area. Machine learning and deep learning algorithms have been implemented in several drug discovery processes such as peptide synthesis, structure-based virtual screening, ligand-based virtual screening, toxicity prediction, drug monitoring and release, pharmacophore modeling, quantitative structure-activity relationship, drug repositioning, polypharmacology, and physiochemical activity. Evidence from the past strengthens the implementation of artificial intelligence and deep learning in this field. Moreover, novel data mining, curation, and management techniques provided critical support to recently developed modeling algorithms. In summary, artificial intelligence and deep learning advancements provide an excellent opportunity for rational drug design and discovery process, which will eventually impact mankind. The primary concern associated with drug design and development is time consumption and production cost. Further, inefficiency, inaccurate target delivery, and inappropriate dosage are other hurdles that inhibit the process of drug delivery and development. With advancements in technology, computer-aided drug design integrating artificial intelligence algorithms can eliminate the challenges and hurdles of traditional drug design and development. Artificial intelligence is referred to as superset comprising machine learning, whereas machine learning comprises supervised learning, unsupervised learning, and reinforcement learning. Further, deep learning, a subset of machine learning, has been extensively implemented in drug design and development. The artificial neural network, deep neural network, support vector machines, classification and regression, generative adversarial networks, symbolic learning, and meta-learning are examples of the algorithms applied to the drug design and discovery process. Artificial intelligence has been applied to different areas of drug design and development process,

such as from peptide synthesis to molecule design, virtual screening to molecular docking, quantitative structure-activity relationship to drug repositioning, protein misfolding to protein-protein interactions, and molecular pathway identification to polypharmacology. Artificial intelligence principles have been applied to the classification of active and inactive, monitoring drug release, pre-clinical and clinical development, primary and secondary drug screening, biomarker development, pharmaceutical manufacturing, bioactivity identification and physiochemical properties, prediction of toxicity, and identification of mode of action.

Keywords: Artificial intelligence, Machine learning, Deep learning, Virtual screening, Drug design and discovery, Artificial neural networks, Computer-aided drug design, Quantitative structure–activity relationship, Drug repurposing

Graphic abstract

The primary concern associated with drug design and development is time consumption and production cost. Further, inefficiency, inaccurate target delivery, and inappropriate dosage are other hurdles that inhibit the process of drug delivery and development. With advancements in technology, computer-aided drug design integrating artificial intelligence algorithms can eliminate the challenges and hurdles of traditional drug design and development. Artificial intelligence is referred to as superset comprising machine learning, whereas machine learning comprises supervised learning, unsupervised learning, and reinforcement learning. Further, deep learning, a subset of machine learning, has been extensively implemented in drug design and development. The artificial neural network, deep neural network, support vector machines, classification and regression, gene



INTRODUCTION

AI is also referred to as machine intelligence, means the ability of computer systems to learn from input or past data. The term AI is commonly used when a machine mimics cognitive behavior associated with the human brain during learning and problem solving . Nowadays, biological and

chemical scientists extensively incorporate AI algorithms in drug designing and discovery process [8]. Computational modeling based on AI and ML principles provides a great avenue for identification and validation of chemical compounds, target identification, peptide synthesis, evaluation of drug toxicity and physiochemical properties, drug monitoring, drug efficacy and effectiveness, and drug repositioning [9]. With the advent of AI principles along with ML and DL algorithms, VS of compounds from chemical libraries, which comprises more than 106 million compounds, become easy and time-effective. Further, AI models eliminate the toxicity problems, which arise due to off-target interactions. Herein, we briefly discuss the evolution of AI from ML to DL and big data involvement in revolutionizing the drug discovery process. Later on, we presented an overview on the congregation of AI and conventional chemistry in the improvement of the drug discovery process and the application of AI in the improvement of the traditional drug discovery process. Afterward, we discuss the numerous AI applications throughout the drug design and discovery processes such as primary and secondary screening, drug toxicity, drug release and monitoring, drug dosage effectiveness and efficacy, drug repositioning, and polypharmacology, and drug-target interactions.

Artificial intelligence (AI) and machine learning (ML) have flourished in the past decade, driven by revolutionary advances in computational technology. This has led to transformative improvements in the ability to collect and process large volumes of data. Meanwhile, the cost of bringing new drugs to market and to patients has become prohibitively expensive. In the remainder of this paper, we use “R&D” to generally describe the research, science, and processes associated with drug development, starting with drug discovery to clinical development and conduct, and finally the life-cycle management stage. Developing a new drug is a long and expensive process with a low success rate as evidenced by the following estimates: average R&D investment is $1.3 billion per drug median development time for each drug ranges from 5.9 to 7.2 years for non-oncology and 13.1 years for oncology; and proportion of all drug-development programs that eventually lead to approval is 13.8% . Recognizing these headwinds, AI/ML techniques are appealing to the drug-development industry, due to their automated nature, predictive capabilities, and the consequent expected increase in efficiency. There is clearly a need, from a patient and a business perspective, to make drug development more efficient and thereby reduce cost, shorten

the development time and increase the probability of success (POS). ML methods have been used in drug discovery for the past 15–20 years with increasing sophistication. The most recent aspect of drug development where a positive disruption from AI/ML is starting to occur, is in clinical trial design, operations, and analysis.

Artificial intelligence (AI) and machine learning (ML) have revolutionized drug development by expediting processes and enhancing decision-making. These technologies analyze vast datasets to identify patterns, predict potential drug candidates, and optimize clinical trial designs. AI's ability to uncover insights from genomics, proteomics, and other biological data accelerates drug discovery, offering a more efficient and targeted approach. ML algorithms also assist in predicting drug interactions, toxicity, and patient response, ultimately reducing costs and time associated with traditional drug development methods.

HISTORY OF ARTIFICIAL INTELLIGENCE OF HEALTH CARE

Artificial intelligence (AI) was first described in 1950; however, several limitations in early models prevented widespread acceptance and application to medicine. In the early 2000s, many of these limitations were overcome by the advent of deep learning. Now that AI systems are capable of analyzing complex algorithms and self-learning, we enter a new age in medicine where AI can be applied to clinical practice through risk assessment models, improving diagnostic accuracy and workflow efficiency. This article presents a brief historical perspective on the evolution of AI over the last several decades and the introduction and development of AI in medicine in recent years. A brief summary of the major applications of AI in gastroenterology and endoscopy are also presented, which are reviewed in further detail by several other articles in this issue of Gastrointestinal Endoscopy.

History of artificial intelligence in healthcare: the first breakthrough of artificial intelligence in healthcare comes in 1950 with the development of turning tests. Later on, in 1975, the first research resource on computers in medicines was developed, followed by NIH's first central AIM workshop marked the importance of artificial intelligence in healthcare. With the development of deep learning in the 2000s and the introduction of Deep QA in 2007, the scope of artificial intelligence in healthcare has increased. Further, in 2010 CAD was applied to endoscopy for the first time, whereas, in 2015, the first Pharmbot was developed. In 2017, the first FDA-approved cloud-based DL application

was introduced, which also marked the implementation of artificial intelligence in healthcare. From 2018 to 2020 several AI trials in gastroenterology were performed.

Classification of artificial intelligence: there are seven classifications of artificial intelligence, which are reasoning and problem solving, knowledge representation, planning and social intelligence, perception, machine learning, robotics: motion and manipulation, and natural language processing, as discussed by Russel and Norvig in their book “Artificial Intelligence: A Modern Approach.” Machine learning is further divided into three significant subsets: supervised learning, unsupervised learning, and deep learning, whereas vision is divided into two subsets, such as image recognition and machine vision. Similarly, speech is divided into two subsets: speech to text and text to speech, whereas natural language processing is classified into five main subsets, including classification, machine translation, question answering, text generation, and content extraction.

Artificial intelligence in the healthcare and pharmaceutical industry has five significant applications, which change the entire scenario. These applications include research and discovery, clinical development, manufacturing and supply chain, patient surveillance, and post-market surveillance.

In 1999, a graphic processing unit (GPU) was launched as a microprocessor circuit, which was developed initially to accelerate 3D graphics processing for computer gaming. Later on, GPUs became popular in the field of technology and research as well because of their ability of parallel computing. A research report presented by META Group in 2001 stated that volume, speed, source

and types of data were increasing, which was a call to prepare for the attack of Big Data. In 2007 Nvidia introduced compute unified device architecture (CUDA), a framework that allowed programmers and researchers to use GPU for general purpose computing [31]. Since then, with the help of CUDA, researchers started using GPUs for DL-driven operations, as high memory bandwidth of GPUs allowed easy handling of massive data involved in DL algorithms, and thousands of cores in GPUs allowed simultaneous parallel processing of neural networks. In 2009, Fei-Fei Li launched ImageNet, which is a free database containing millions of labeled images that can be used for research purposes. AlexNet, a convolutional neural network, was created by Alex Krizhevsky around 2012, which helped in strengthening the speed and dropout using rectified linear units . In the same year, "the cat experiment" conducted by Google Brain concluded that the network correctly recognizes less than 16% of the presented objects . In 2014 Nvidia introduced CUDA deep neural network (cuDNN), a CUDA-based DL library, which accelerated DL-based operations . Similarly, "Deep Face" was developed and released in 2014 to identify faces with 97.5% accuracy . In the same year, generative adversarial networks (GANs) were introduced, using two competing neural networks to check whether the data are genuine or generated.In 2016, Cray Inc. used Microsoft's neural network software on its XC50 supercomputer with 1000 Nvidia Tesla P100 GPUs that could perform the task and gave output in a fraction of seconds. In 2017 Nvidia introduced Tesla V100 GPU, which had tensor cores that accelerated AI-based operations. However, DL is still in its growth phase, and creative ideas are required for further advancement in this field.

FOUNDATION OF ARTIFICIAL INTELLIGENCE

Artificial intelligence, or AI, refers to the simulation of human intelligence by software-coded heuristics. Nowadays this code is prevalent in everything from cloud-based, enterprise applications to consumer apps and even embedded firmware.

The ideal characteristic of artificial intelligence is its ability to rationalize and take actions that have the best chance of achieving a specific goal. A subset of artificial intelligence is machine learning (ML), which refers to the concept that computer programs can automatically learn from and adapt to new data without being assisted by humans. Deep learning techniques enable this automatic learning through the absorption of huge amounts of unstructured data such as text, images, or video.

Artificial Intelligence (AI) has recently started to gear-up its application in various sectors of the society with the pharmaceutical industry as a front-runner beneficiary. This review highlights the impactful use of AI in diverse areas of the pharmaceutical sectors viz., drug discovery and development, drug repurposing, improving pharmaceutical productivity, clinical trials, etc. to name a few, thus reducing the human workload as well as achieving targets in a short period. Crosstalk on the tools and techniques utilized in enforcing AI, ongoing challenges, and ways to overcome them, along with the future of AI in the pharmaceutical industry, is also discussed.

In drug discovery, AI is relevant in several ways:

1. *Data Analysis:* AI algorithms can analyze vast datasets, including genomic and proteomic information, to identify patterns and potential drug targets.Drug discovery data analysis involves the application of computational methods to analyze large and complex datasets generated during the drug discovery process. This analysis plays a crucial role in identifying potential drug candidates, optimizing drug design, and predicting drug efficacy and safety. From a business perspective, drug discovery data analysis offers several key benefits and applications

2. *Target Identification:* AI helps identify specific biological targets implicated in diseases, aiding in the selection of proteins or genes that could be targeted for therapeutic intervention. It is the process of identifying the direct molecular targets of small molecules, such as nucleic acid and proteins. Target identification can be approached by computational methods, direct biochemical methods, and genetic interactions.

3. *Compound Screening:* AI accelerates the screening of chemical compounds by predicting their potential biological activity, reducing the time and resources needed for experimental testing.Compound Screening is defined as the identification of compounds that could be promising candidates for drug development, before it advances to the more-costly stages of preclinical and clinical trials. Compound Screening is also done to identify the potential side effects caused by administration of such compound(s).

4. *Drug Design:* Generative models within AI can assist in designing novel drug candidates by predicting molecular structures with desired properties, optimizing for efficacy and safety.Drug design is the inventive process of finding new medications based on the knowledge of a biological target. In the most basic sense, drug design involves the design of molecules that are complementary in shape and charge to the molecular target with which they interact and bind.

Revolutionizing drug discovery process: role of big data and artificial intelligence

Big data can be defined as data sets that are too gigantic and intricate to be analyzed with the conventional data analyzing software, tools, and techniques. The three main characteristic features of big data are volume, velocity, and variety, where volume represents the huge amount and mass of data generated, velocity represents the rate at which these data are being reproduced, and variety represents heterogenicity present in the data sets. With the advent of microarray, RNA-seq, and high-throughput sequencing (HTS) technologies, a plethora of biomedical data is being engendered every day, due to which contemporary drug discovery has made a transition into the big data era. In drug discovery, the first and foremost step is the identification of appropriate targets (e.g., genes, proteins) involved in disease pathophysiology, followed by finding suitable drugs or drug-like molecules which can meddle with these targets, and now we have access to a constellation of biomedical data repositories which can help us in this regard . Moreover, the evolution of AI has made big data analytics a lot easier as there is a myriad of ML techniques available now, which can help in extracting useful features, patterns, and structures present in these big biomedical data sets . For target identification, a feature like a gene expression is widely used to understand disease mechanisms and find genes responsible for the disease. Microarray and

RNA-seq technologies have generated a large amount of gene expression data for various disorders. NCBI Gene Expressions are some of the big repositories which contain gene expression data. By analyzing gene expression signatures, we can find out target genes responsible for different disorders. For example, using the ML approach and gene expression data, van IJzendoorn et al. 2019 found out novel biomarkers and potential drug targets for rare soft tissue sarcoma.

The revolutionizing of the drug discovery process is significantly influenced by the integration of big data and artificial intelligence (AI). Here's how they play pivotal roles:

1. Data Integration: Big data encompasses diverse sources of information, including genomics, proteomics, patient records, and scientific literature. AI algorithms can process and integrate this massive amount of data, revealing complex relationships and patterns that might be challenging for traditional methods to identify.

2. Target Identification: AI analyzes biological data to pinpoint potential drug targets. By understanding the intricacies of diseases at a molecular level, researchers can identify specific proteins or genes that may serve as effective targets for drug development.

3. Predictive Modeling: AI algorithms can predict the potential efficacy and safety of drug candidates, enabling researchers to prioritize the most promising compounds for further testing. This reduces the time and resources required for experimental validation.

4. Drug Design: AI facilitates the design of novel drug candidates by generating molecular structures that meet specific criteria, such as binding affinity and pharmacokinetic properties. This accelerates the drug discovery process by providing a more focused starting point for synthesis.

5. Personalized Medicine: Combining big data and AI allows for a more precise understanding of individual patient characteristics. This leads to the development of personalized treatment plans, considering genetic variations and other factors, improving therapeutic outcomes.

6.High-Throughput Screening: AI enhances the efficiency of high-throughput screening by automating the analysis of large compound libraries. This accelerates the identification of potential drug candidates and streamlines the experimental process.

7.Clinical Trial Optimization: Big data analytics and AI contribute to more efficient clinical trial design and execution. Predictive models can identify suitable patient cohorts, optimize trial protocols, and predict potential outcomes.

APPLICATION OF BIG FOR DRUG DESIGNING DISCOVERY

Application of big data for drug designing and discovery: with the increase in biological and chemical data from the literature, in vitro, in vivo, clinical studies, genomics studies, proteomics studies, metabolomics studies, gene ontology studies, and molecular pathway data, different data repositories have been developed. For instance, ChemSpider, ChEMB, ZINC, BindingDB, and PubChem are the essential databases for compound synthesis and screening in the drug designing and discovery process. The data stored in the above-said databases were curated and screened out

for pharmacological and physicochemical properties of compound necessary for the drug discovery process instead of quantum mechanical calculations such as solvation energy and proton affinity the wave function, atomic forces, and transition state. The high-throughput screened data were subject to filtration based on drug-likeness, PAINS calculation, ADMET analysis, and toxicity. The filtered compounds were subject to artificial intelligence models such as deep learning, random forest, classification and regression, and neural networks for further analysis. These compounds were then subjected to quantitative-structure activity relationship and pharmacophore models followed by molecular docking and molecular dynamics simulations studies. Afterward, the final predicted compounds were visualized for binding energy calculations and active site identification. Thus, the final compound was identified and underwent in vitro and in vivo experimental studies for validation. However, quantum mechanical properties play a crucial role in the process of drug discovery and designing, but these properties cannot directly hamper the process of drug designing. QM methods include ab initio density functional theory and semi-empirical calculations, where accurate calculations use electron correlation methods. QM will become a more prominent tool in the repertoire of the computational medicinal chemist. Therefore, modern QM approaches will play a more direct role in informing and streamlining the drug-discovery process.

Big data plays a crucial role in drug designing and discovery by providing valuable insights and facilitating various aspects of the process. Here are key applications:

1. Target Identification and Validation: Big data analysis helps identify and validate potential drug targets by integrating information from genomics, proteomics, and other biological datasets. This accelerates the selection of promising targets for further investigation.

2. Disease Profiling: Comprehensive analysis of large-scale biological and clinical datasets allows for a better understanding of diseases at the molecular level. This information aids in identifying specific biomarkers and pathways associated with diseases, guiding the development of targeted therapies.

3.Drug Repurposing: Big data enables the exploration of existing datasets to identify drugs that may have therapeutic effects for different conditions. This approach can uncover new uses for existing drugs, potentially expediting the drug development process.

4. Chemoinformatics: Analyzing chemical data on a large scale aids in the prediction of molecular structures and properties. This is particularly useful for designing new drug candidates with desired pharmacological profiles.

5. High-Throughput Screening (HTS): Big data analytics enhances the efficiency of HTS by automating the analysis of large compound libraries. This accelerates the identification of potential drug candidates, streamlining the experimental process.

6. Predictive Modeling: Big data-driven predictive models assess the potential efficacy and safety of drug candidates. This allows researchers to prioritize compounds for further development, reducing the time and resources needed for experimental validation.

7. Clinical Trial Optimization: Analyzing diverse datasets, including patient records and real-world evidence, helps optimize clinical trial design. This includes identifying suitable patient populations,

KEEP DEEP LEARNING TECHNIQUE

Deep learning techniques play a crucial role in drug discovery, offering powerful tools for analyzing complex biological data. Here are key deep learning techniques in this field:

1.Convolutional Neural Networks (CNNs): Used for image analysis, CNNs can process and analyze molecular and cellular images, aiding in the identification of patterns and structures relevant to drug discovery. An architecture commonly used in Computer Vision. Computer vision is a field of Artificial Intelligence that enables a computer to understand and interpret the image or visual data.

When it comes to Machine Learning, Artificial Neural Networks perform really well. Neural Networks are used in various datasets like images, audio, and text. Different types of Neural Networks are used for different purposes, for example for predicting the sequence of words we use Recurrent Neural Networks more precisely an LSTM, similarly for image classification we use Convolution Neural networks. In this blog, we are going to build a basic building block for CNN

2.Recurrent Neural Networks (RNNs): Particularly useful for sequential data, RNNs can model temporal dependencies in biological sequences, such as DNA, RNA, or protein sequences. This helps in predicting molecular interactions and understanding complex biological processes.

A Deep Learning approach for modelling sequential data is Recurrent Neural Networks (RNN). RNNs were the standard suggestion for working with sequential data before the advent of attention models. Specific parameters for each element of the sequence may be required by a deep feedforward model. It may also be unable to generalize to variable-length sequences.

Recurrent Neural networks

Recurrent Neural Networks use the same weights for each element of the sequence, decreasing the number of parameters and allowing the model to generalize to sequences of varying lengths. RNNs generalize to structured data other than sequential data, such as geographical or graphical data, because of its design.

Recurrent neural networks, like many other deep learning techniques, are relatively old. They were first developed in the 1980s, but we didn't appreciate their full potential until lately. The advent of long short-term memory (LSTM) in the 1990s, combined with an increase in computational power and the vast amounts of data that we now have to deal with, has really pushed RNNs to the forefront.

3.Generative Adversarial Networks (GANs): GANs can generate new, realistic molecular structures. In drug discovery, GANs assist in the generation of novel chemical compounds with desired properties, aiding in the exploration of diverse chemical space.

Generative modeling is an unsupervised learning task in machine learning that involves automatically discovering and learning the regularities or patterns in input data in such a way that the model can be used to generate or output new examples that plausibly could have been drawn from the original dataset.

GANs are a clever way of training a generative model by framing the problem as a supervised learning problem with two sub-models: the generator model that we train to generate new examples, and the discriminator model that tries to classify examples as either real (from the domain) or fake (generated). The two models are trained together in a zero-sum game, adversarial, until the discriminator model is fooled about half the time, meaning the generator model is generating plausible examples.

GANs are an exciting and rapidly changing field, delivering on the promise of generative models in their ability to generate realistic examples across a range of problem domains, most notably in image-to-image translation tasks such as translating photos of summer to winter or day to night, and in generating photorealistic photos of objects, scenes, and people that even humans cannot tell are fake.

4. Graph Neural Networks: (GN Well-suited for modeling molecular structures, GNNs can represent chemical compounds as graphs, capturing relationships between atoms and bonds.

Artificial intelligence in primary and secondary drug screening

Today AI has come out as a very successful and demanding technology because it saves time and is cost-efficient [93]. In general, cell classification, cell sorting, calculating properties of small molecules, synthesizing organic compounds with the help of computer programs, designing new compounds, developing assays, and predicting the 3D structure of target molecules are some time-consuming and tiresome tasks which with the help of AI can be reduced and can speed up the process of drug discovery. The primary drug screening includes the classification and sorting of cells by image analysis through AI technology. Many ML models using different algorithms recognize images with great accuracy but become incompetent when analyzing big data. To classify the target cell, firstly, the ML model needs to be trained so that it can identify the cell and its features, which is basically done by contrasting the image of the targeted cells, which separates it from the background. Images with varying textured features like wavelet-based texture features and Tamura texture features are extracted, which is further reduced in dimensions through principal component analysis (PCA). A study suggests that least-square SVM (LS-SVM) showed the highest classification accuracy of 95.34%. Regarding cell sorting, the machine needs to be fast to separate out the targeted cell type from the given sample. Evidence suggests that image-activated cell sorting (IACS) is the most advanced device that could measure the optical, electrical, and mechanical properties of the cell.

Artificial intelligence in primary and secondary drug screening: in drug discovery and designing pipeline, screening of potential lead is crucial, and artificial intelligence plays a great role in identifying novel and potential lead compounds. There are approximately 106 million chemical structure presents in chemical space from different studies such as OMIC studies, clinical and pre-clinical studies, in vivo assays, and microarray analysis. With machine learning models such as reinforcement models, logistic models, regression models, and generative models, these chemical structures are screened out based on active sites, structure, and target binding ability. The complete drug discovery process through artificial intelligence will take about 14–18 years, which is comparatively less than the traditional drug discovery process. The first step in the drug discovery process is lead identification, in which disease-modifying target protein is identified through reverse docking, bioinformatics analysis, and computational chemical biology. In the second step, primary screening of compounds is done to select potential lead compounds, which can inhibit target protein. This can be done through virtual screening and de novo designing. The next step in the drug discovery process includes lead optimization and lead compound identification through focused library design, drug-like analysis, drug-target reproducibility, and computational biology. Afterward, secondary screening of compounds is performed, followed by pre-clinical trials. The drug discovery process's final step is clinical development through cell-culture analysis, animal model experimentation, and patient analysis.

The secondary drug screening includes analyzing the physical properties, bioactivity, and toxicity of the compound. Melting point and partition coefficient are some of the physical properties that govern the compound's bioavailability and are also essential to design new compounds, while designing a drug, molecular representation can be done using different methods like molecular fingerprinting, simplified molecular-input line-entry system (SMILES), and Coulomb matrices. These data can be used in DNN, which comprises two different stages, namely generative and predictive stage. Though both the stages are trained separately through supervised learning, when they are trained jointly, bias can be applied to the output.

Applications of artificial intelligence in drug development process

The most arduous and desponding step in the drug discovery and development process is identifying suitable and bioactive drug molecules present in the vast size of chemical space, which is in the order of 1060 molecules. Further, the drug discovery and development process are considered a time- and cost-consuming process. The most infuriating point is that nine out of ten drug molecules usually fail to pass phase II clinical trials and other regulatory approvals. The above-said limitations of drug discovery and development can be addressed by implementing AI-based tools and techniques. AI is involved in every stage of the drug development process such as small molecules design, identification of drug dosage and associated effectiveness, prediction of bioactive agents, protein–protein interactions, identification of protein folding and misfolding, structure and ligand-based VS, QSAR modeling, drug repurposing, prediction of toxicity and bioactive properties, and identification of mode of action of drug compounds as discussed below.

Peptide synthesis and small molecule design

Peptide are a biologically active small chain of around 2–50 amino acids, which are increasingly being explored for therapeutic purposes as they have the ability to cross the cellular barrier and can reach the desired target site . In recent years, researchers have taken advantage of AI and used it to discover novel peptides. For instance, Yan et al. 2020 developed Deep-AmPEP30, a DL-based platform for the identification of short anti-microbial peptides (AMPs) . Deep-AmPEP30 is a CNN-driven tool that predicts short AMPs from DNA sequence data. Using Deep-AmPEP30, Yan et al. identified novel AMPs from the genome sequence of C. glabrate, a fungal pathogen present in the GI tract. Likewise, Plisson et al. 2020 combined the ML algorithm with an outlier detection technique to discover AMPs with non-hemolytic profiles . In addition, Kavousi et al. developed IAMPE , a web server for the identification of anti-microbial peptides, which integrates 13CNMR-based features and physicochemical features of peptides as input to ML algorithms, in order to identify novel AMPs.

Moreover, small molecules are molecules that have very low molecular weight, and like peptides, small molecules are too being explored for therapeutic purposes using AI-based tools. For instance, Zhavoronkov et al. The devised generative tensorial reinforcement learning (GENTRL), a generative reinforcement learning-based tool for the de novo design of small molecules.

Identification of drug dosage and drug delivery effectiveness

Administering an improper dose of any drug to a patient can lead to undesirable and lethal side effects; hence, it is crucial to determine a safe drug dose for treatment purposes. Over the years, it has been challenging to ascertain the optimum dose of a drug that can achieve the desired efficacy with minimum toxic side effects. With the emergence of AI, lots of researchers are taking the help of ML and DL algorithms to determine appropriate drug dosage. For instance, Shen et al. developed an AI-based platform, referred to as AI-PRS, to determine the optimum dose and combinations of drugs to be used for HIV treatment through antiretroviral therapy. AI-PRS is a neural network-driven approach, which relates drug combinations and dosage to efficacy through a parabolic response curve (PRS). In their study, Shen et al. administered a combination of tenofovir, efavirenz, and lamivudine to 10 HIV patients, and in due course, using the PRS method, they found out the dose of tenofovir could be reduced by 33% of the starting dose without causing virus relapse. Hence, using AI-PRS optimum drug dosage can be found out for other diseases as well. Further, Pantuck et al. [123] developed CURATE.AI, to determine adequate drug dose, which uses a patient's personal data and transforms it to CURATE.AI profile in order to ascertain optimum dose.

Future challenges and possible solutions

At present, the major challenge for the pharmaceutical industry while developing a new drug is its increased costs and reduced efficiency. However, ML approaches and recent developments in DL come with great opportunities to reduce this cost, increase efficiency, and save time during the drug discovery and development process. Advances in AI algorithms, especially in DL approaches along with improving architectural hardware and easy accessibility of big data, are all indicating toward the third wave of AI. AI approaches in drug development have aroused great interest among researchers, such that many pharmaceutical companies have collaborated with AI companies.

Moreover, the number of startups in this field has also escalated and reached 230 by June 2020 . Further, DL approaches integrate data at multiple levels through nonlinear models, which is the shortcoming of the AI and ML approaches. However, integration of data at multiple levels makes DL algorithm advantageous as it provides great accuracy and precision. Moreover, in comparison with AI and ML algorithms, DL provides a much more flexible architecture to create a neural network for a specific problem. Applications of AI like natural language processing, image, and voice recognition are easily doable these days, which has beaten humans in terms of performance. So, it comes with no surprise that AI can very well be used in the drug discovery process. Today, AI is used in drug discovery for target identification, hit discovery, lead optimization, ADMET prediction, and structuring clinical trials. Despite great success, there are many remaining challenges like high-quality data acquisition under which there are two significant concerns. Firstly, labeling cannot be binary as the action of drugs in biological systems is complicated; secondly, the amount of data available in drug discovery is infinitesimal compared to the enormous amount of information available. Therefore, a community is required that not only provides quantity but the quality of data. In the pharmaceutical industry, open data sharing is not common, and Pistoia alliance has taken the initiative to start a movement that has encouraged many companies to share their data with others. They also intend to establish a uniform data format, which is technically challenging. A possible solution to deal with this problem is to develop an algorithm that can handle sparse data; one such has been developed by Stanford University named "one-shot learning," which predicts properties of a drug on the basis of heterogeneous data. Moreover, the accuracy and uncertainty of the experimental data can be used for model building, that is instead of establishing new ML technologies, one can put efforts in training the existing one by tuning large number of hyperparameters and optimizing it for good results, although some studies indicated that some reasonable parameters can be used to start the optimization. Molecular representation is also a challenge as it is one of the governing factors in model building. Few recently developed models learn task-related features from the raw data and refine the molecular representation to a standard. Earlier, drug repurposing used to rely only on clinical observations. However, the current large amount of data comprising of scientific literature, patents, and clinical trial results can collectively be used to improve the screening process.

Additionally, DL-based VS can make full use of the data and reduce false-positive rates obtained due to imbalance in positive and negative data. Lead optimization is also a challenge in order to develop an efficient drug with good ADMET properties and target activities; however, these parameters are independent and at times mutually incompatible with each other. This problem can be solved by optimizing each parameter separately and further improving the model. Pharmaceutical companies' faces trouble recruiting sufficient number of patients for clinical trials.

Ethical Considerations Regarding the Use of AI in the drug discovery

As discussed in the previous section, it is important to consider the ethical implications of using AI in this field. One key issue is the potential for AI to be used to make decisions that affect people's health and well-being, such as decisions about which drugs to develop, which clinical trials to conduct, and how to market and distribute drugs. Another key concern is the potential for bias in AI algorithms, which could result in unequal access to medical treatment and the unfair

treatment of certain groups of people. This could undermine the principles of equality and justice. The use of AI in the pharmaceutical industry also raises concerns about job losses due to automation. It is important to consider the potential impact on workers and provide support for those who may be affected. Additionally, the use of AI in the pharmaceutical industry raises questions about data privacy and security. As AI systems rely on large amounts of data to function, there is a risk that sensitive personal information could be accessed or misused. This could have serious consequences for individuals, as well as for the reputation of the companies involved. The collection and use of sensitive medical data must be performed in a way that respects the individuals' privacy and complies with the relevant regulations.

Overall, the ethical use of AI in the pharmaceutical industry requires careful consideration and the adoption of thoughtful approaches to addressing these concerns. This can include measures such as ensuring that AI systems are trained on diverse and representative data, regularly reviewing and auditing AI systems for bias, and implementing strong data privacy and security protocols. By addressing these issues, the pharmaceutical industry can use AI in a responsible and ethical manner.

Conclusions and Summary of the Potential of AI for Revolutionizing Drug Discovery

In conclusion, AI has the potential to revolutionize the drug discovery process, offering improved efficiency and accuracy, accelerated drug development, and the capacity for the development of more effective and personalized treatments. However, the successful application of AI in drug discovery is dependent on the availability of high-quality data, the addressing of ethical concerns, and the recognition of the limitations of AI-based approaches.

Graphical flowchart illustrating the development process of a pharmacologically active molecule, from design to knowledge communication and transfer. AI-based approaches complement traditional methods but still cannot replace human expertise. By combining the predictive power of AI with human researchers' knowledge, the drug discovery process can be optimized and accelerated. The present work examines the cutting-edge advancements in the stages of "Literature revision and analysis" and "Write scientific reports and publications" (highlighted in red) a chatbot based on the GPT-3.5 language model.

Recent developments in AI, including the use of data augmentation, explainable AI, and the integration of AI with traditional experimental methods, offer promising strategies for overcoming the challenges and limitations of AI in the context of drug discovery. The growing levels of interest and attention from researchers, pharmaceutical companies, and regulatory agencies, combined with the potential benefits of AI, make this an exciting and promising area of research, with the potential to transform the drug discovery process.

REFERENCE

1. Lipinski CF, Maltarollo VG, Oliveira PR, et al. Advances and perspectives in applying deep learning for drug design and discovery. Front Robot AI. 2019 doi: 10.3389/frobt.2019.00108. [PMC free article] [PubMed] [CrossRef] [Google Scholar]

2. Hamet P, Tremblay J. Artificial intelligence in medicine. Metabolism. 2017 doi: 10.1016/j.metabol.2017.01.011. [PubMed] [CrossRef] [Google Scholar]

3. Hassanzadeh P, Atyabi F, Dinarvand R. The significance of artificial intelligence in drug delivery system design. Adv Drug Deliv Rev. 2019 doi: 10.1016/j.addr.2019.05.001. [PubMed] [CrossRef] [Google Scholar]

4. Duch W, Swaminathan K, Meller J. Artificial intelligence approaches for rational drug design and discovery. Curr Pharm Des. 2007 doi: 10.2174/138161207780765954. [PubMed] [CrossRef] [Google Scholar]

5. Zhang L, Tan J, Han D, Zhu H. From machine learning to deep learning: progress in machine intelligence for rational drug discovery. Drug Discov Today. 2017 doi: 10.1016/j.drudis.2017.08.010. [PubMed] [CrossRef] [Google Scholar]

6. 6. Jordan AM. Artificial intelligence in drug design–the storm before the calm? ACS Med Chem Lett. 2018 doi: 10.1021/acsmedchemlett.8b00500. [PMC free article] [PubMed] [CrossRef] [Google Scholar]

7. Goel AK, Davies J (2019) Artificial intelligence. In: The Cambridge Handbook of Intelligence. Cambridge

8. Harrer S, Shah P, Antony B, Hu J. Artificial Intelligence for Clinical Trial Design. Trends Pharmacol: Sci; 2019. [PubMed] [Google Scholar]

9. 9. Zhong F, Xing J, Li X, et al. Artificial intelligence in drug design. Sci China Life Sci. 2018 doi: 10.1007/s11427-018-9342-2. [PubMed] [CrossRef] [Google Scholar]

10. Brown N, Ertl P, Lewis R, et al. Artificial intelligence in chemistry and drug design. J Comput Aided Mol Des. 2020 doi: 10.1007/s10822-020-00317-x. [PubMed] [CrossRef] [Google Scholar]

11. Badillo S, Banfai B, Birzele F, et al. An introduction to machine learning. Clin Pharmacol Ther. 2020 doi: 10.1002/cpt.1796. [PMC free article] [PubMed] [CrossRef] [Google Scholar]

12. Dutta Majumdar D. Trends in pattern recognition and machine learning. Def Sci J. 1985 doi: 10.14429/dsj.35.6027. [CrossRef] [Google Scholar]

13. 13. Kubat M (2017) An Introduction to Machine Learning

14. Aggarwal M, Murty MN (2021) Deep Learning. In: SpringerBriefs in Applied Sciences and Technology. 10.1007/978-981-33-4022-0_3

15. Schmidhuber J. Deep learning in neural networks: an overview. Neural Netw. 2015 doi: 10.1016/j.neunet.2014.09.003. [PubMed] [CrossRef] [Google Scholar]

16. Hu YH, Hwang JN (2001) Introduction to neural networks for signal processing. In: Handbook of Neural Network Signal Processing. CRC Press, pp 12–41

17. Angermueller C, Pärnamaa T, Parts L, Stegle O. Deep learning for computational biology. Mol Syst Biol. 2016 doi: 10.15252/msb.20156651. [PMC free article] [PubMed] [CrossRef] [Google Scholar]

18. McCulloch WS, Pitts W. A logical calculus of the ideas immanent in nervous activity. Bull Math Biophys. 1943;5:115–133. doi: 10.1007/BF02478259. [PubMed] [CrossRef] [Google Scholar]

19. Turing AM. Parsing the Turing Test: Philosophical and Methodological Issues in the Quest for the Thinking Computer. Netherlands: Springer; 2009. Computing machinery and intelligence; pp. 23–65. [Google Scholar]

20. Samuel AL. Some studies in machine learning using the game of checkers. IBM J Res Dev. 1959;3:210–229. doi: 10.1147/rd.33.0210. [CrossRef] [Google Scholar]

21. 21. Rosenblatt F (1957) The Perceptron: A Perceiving and Recognizing Automaton, Report 85–60–1

22. KELLEY HJ, Gradient theory of optimal flight paths. ARS J. 1960;30:947–954. doi: 10.2514/8.5282. [CrossRef] [Google Scholar]

23. Dreyfus S. The numerical solution of variational problems. J Math Anal Appl. 1962;5:30–45. doi: 10.1016/0022-247X(62)90004-5. [CrossRef] [Google Scholar]

24. Fukushima K. Neocognitron: a self-organizing neural network model for a mechanism of pattern recognition unaffected by shift in position. Biol Cybern. 1980;36:193–202. doi: 10.1007/BF00344251. [PubMed] [CrossRef] [Google Scholar]

25. Tripathy RK, Mahanta S, Paul S. Artificial intelligence-based classification of breast cancer using cellular images. RSC Adv. 2014;4:9349–9355. doi: 10.1039/c3ra47489e. [CrossRef] [Google Scholar]

26. Samui P, Kothari DP. Utilization of a least square support vector machine (LSSVM) for slope stability analysis. Sci Iran. 2011;18:53–58. doi: 10.1016/j.scient.2011.03.007. [CrossRef] [Google Scholar]

27. Chan HCS, Shan H, Dahoun T, et al. Advancing Drug Discovery via Artificial Intelligence. Trends Pharmacol Sci. 2019;40:592–604. doi: 10.1016/j.tips.2019.06.004. [PubMed] [CrossRef] [Google Scholar]

28. Ho CWL, Soon D, Caals K, Kapur J. Governance of automated image analysis and artificial intelligence analytics in healthcare. Clin Radiol. 2019;74:329–337. doi: 10.1016/j.crad.2019.02.005. [PubMed] [CrossRef] [Google Scholar]

29. Andrysek T (2003) Impact of physical properties of formulations on bioavailability of active substance: Current and novel drugs with cyclosporine. In: Molecular Immunology; 39(17–18):1061–5. 10.1016/s0161-5890(03)00077-4. [PubMed]

30. Elton DC, Boukouvalas Z, Butrico MS, et al. Applying machine learning techniques to predict the properties of energetic materials. Sci Rep. 2018;8:9059. doi: 10.1038/s41598-018-27344-x. [PMC free article] [PubMed] [CrossRef] [Google Scholar]

31. Tyrchan C, Evertsson E. Matched molecular pair analysis in short: algorithms, applications and limitations. Comput Struct Biotechnol J. 2017;15:86–90. doi: 10.1016/j.csbj.2016.12.003. [PMC free article] [PubMed] [CrossRef] [Google Scholar]

32. Turk S, Merget B, Rippmann F, Fulle S. Coupling matched molecular pairs with machine learning for virtual compound optimization. J Chem Inf Model. 2017;57:3079–3085. doi: 10.1021/acs.jcim.7b00298. [PubMed] [CrossRef] [Google Scholar]

33. Turk S, Merget B, Rippmann F, Fulle S. Coupling matched molecular pairs with machine learning for virtual compound optimization. J Chem Inf Model. 2017;57:3079–3085. doi: 10.1021/acs.jcim.7b00298. [PubMed] [CrossRef] [Google Scholar]

34. Carpenter KA, Huang X. Machine learning-based virtual screening and its applications to Alzheimer's drug discovery: a review. Curr Pharm Des. 2018;24:3347–3358. doi: 10.2174/1381612824666180607124038. [PMC free article] [PubMed] [CrossRef] [Google Scholar]

35. Schyman P, Liu R, Desai V, Wallqvist A. vNN web server for ADMET predictions. Front Pharmacol. 2017;8:889. doi: 10.3389/fphar.2017.00889. [PMC free article] [PubMed] [CrossRef] [Google Scholar]

36. Álvarez-Machancoses Ó, Fernández-Martínez JL. Using artificial intelligence methods to speed up drug discovery. Expert Opin Drug Discov. 2019;14(8):769–777. doi: 10.1080/17460441.2019.1621284. [PubMed] [CrossRef] [Google Scholar]

37. Fleming N. How artificial intelligence is changing drug discovery. Nature. 2018 doi: 10.1038/d41586-018-05267-x. [PubMed] [CrossRef] [Google Scholar]

38. Segler MHS, Kogej T, Tyrchan C, Waller MP. Generating focused molecule libraries for drug discovery with recurrent neural networks. ACS Cent Sci. 2018 doi: 10.1021/acscentsci.7b00512. [PMC free article] [PubMed] [CrossRef] [Google Scholar]

39. 53. Jarrahi M.H. Artificial intelligence and the future of work: Human-AI symbiosis in organizational decision making. Bus. Horiz. 2018;61:577–586. doi: 10.1016/j.bushor.2018.03.007. [CrossRef] [Google Scholar]

40. Wang L., Ding J., Pan L., Cao D., Jiang H., Ding X. Artificial intelligence facilitates drug design in the big data era. Chemom. Intell. Lab. Syst. 2019;194:103850. doi: 10.1016/j.chemolab.2019.103850. [CrossRef] [Google Scholar]

41. Naik N., Hameed B.M.Z., Shetty D.K., Swain D., Shah M., Paul R., Aggarwal K., Brahim S., Patil V., Smriti K., et al. Legal and Ethical Consideration in Artificial Intelligence in Healthcare: Who Takes Responsibility? Front. Surg. 2022;9:266. doi: 10.3389/fsurg.2022.862322. [PMC free article] [PubMed] [CrossRef] [Google Scholar]