ARTIFICIAL INTELLIGENCE AND MACHINE LEARNING BASED NEW DRUG DISCOVERY PROCESS WITH MOLECULAR MODELING

Author : Laxmi Sushant Sarode.

Author Affiliation : PRES College of Pharmacy for (womens) Chincholi , Sinnar , Nashik.

ABSTRACT

Artificial intelligence is a branch of computer science that built smart machines, which are capable to perform task that require human intelligence. But artificial intelligence does not confine itself to methods that are biologically observable.

Now a days Artificial Intelligence is widely used technology . AI is used in various sectors such as robotics, healthcare, agriculture, social media,finance and machine learning. Artificial Intelligence is one of the Computer Aided Drug Discovery (CADD) and machine learning is one of its subtopic. It improves drug discovery and development for more efficacy and accuracy.[1]

The current rise of artificial intelligence and machine learning has been significant. It has reduced the human workload improved quality of life significantly. This article describes the use of artificial intelligence and machine learning to augment drug discovery and development to make them more efficient and accurate. In this study, a systematic evaluation of studies was carried out; these were selected based on prior knowledge of the authors and a keyword search in publicly available databases which were filtered based on related context, abstract, methodology, and full text. This body of work supported the roles of machine learning and artificial intelligence in facilitating drug development and discovery processes, making them more cost-effective or altogether eliminating the need for clinical trials, owing to the ability to conduct simulations using these technologies. The results of this paper demonstrate the prevalent application of machine learning and artificial intelligence methods in drug discovery, and indicate a promising future for these technologies; these results should enable researchers, students, and pharmaceutical industry to dive deeper into machine learning and artificial intelligence in a drug discovery and development context.[3]

Keywords : Artificial intelligence ; Machine learning ; Drug discovery ; QSAR ; synthetic prediction; Novel drug discovery ; one shot learning ; CADD ; Big data in medicine.

Main Text :

Every side of life is constantly change, and the humans have to control these changes for our benefit; especially in the field of medicine and pharmaceuticals.

For many decades, the manufacturing of drug products is controlled by a legal field that checks the quality of final products by testing of raw , in-process materials, finished product, batch-based operations and fixed process conditions . .

The drug and biopharmaceutical industries have been limited source of inventive and novel technologies or machinery, and have led the development of novel principles or interpretations in general chemical and mechanical engineering. The pharmaceutical industry is in critical need of mechanical innovation, easing the creation of medications for human use. Creating and manufacturing complex processes medications that are safe for humans on a commercial scale, and incorporating them into mainstream therapeutic use, has been challenging, owing to

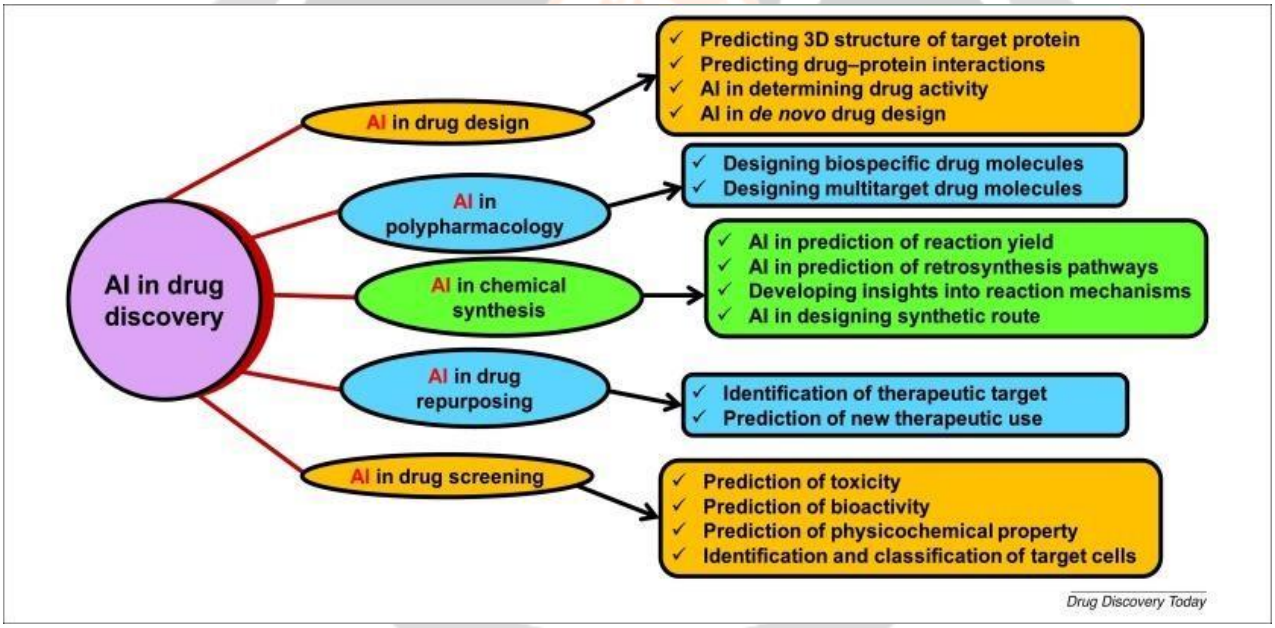
existing limitations on technological resources . The use of artificial intelligence (AI) is increasing, and is likely to change how clinical examination and training is carried out. Doctors can participate in the development of this technology for use in the medical and pharmaceutical industries; this will ensure that the potential of AI to significantly improve medical care is fulfilled[2].

AI is currently used in the pharmaceutical industry in four main ways. The first is in the assessment of the severity of disease and the prediction of whether treatment will be successful for an individual patient, even prior to its administration. Secondly, it is used to prevent or solve complications during treatment. Its third main use is as an assistive technology to during treatment procedures or operations on patients. Lastly, it is used to determine the reasons behind the use of particular instruments or chemicals during treatment, and to develop or extrapolate new uses for instruments or chemicals to improve safety and efficacy.

AI also has a more general role in the management and analysis of big data.[3]

AI in drug discovery :

The vast chemical space, comprising $>10^{60}$ molecules, fosters the development of a large number of drug molecules . However, the lack of advanced technologies limits the drug development process, making it a time-consuming and expensive task, which can be addressed by using AI . AI can recognize hit and lead compounds, and provide a quicker validation of the drug target and optimization of the drug structure design .
Different applications of AI in drug discovery are depicted in Figure 1.



- Download : Download high-res image (808KB)
- Download : Download full-size image

Figure 1. Role of artificial intelligence (AI) in drug discovery. AI can be used effectively in different parts of drug discovery, including drug design, chemical synthesis, drug screening, polypharmacology, and drug repurposing.

Despite its advantages, AI faces some significant data challenges, such as the scale, growth, diversity, and uncertainty of the data. The data sets available for drug development in pharmaceutical companies can involve millions of compounds, and traditional ML tools might not be able to deal with these types of data. Quantitative structure-activity relationship (QSAR)-based computational model can quickly predict large numbers of compounds or simple physicochemical parameters, such as log P or log D. However, these models are some way from the predictions of complex biological properties, such as the efficacy and adverse effects of compounds. In addition, QSAR- based models also face problems such as small training sets, experimental data error in training sets, and lack of experimental validations. To overcome these challenges, recently developed AI approaches, such as DL and relevant modeling studies, can be implemented for safety and efficacy evaluations of drug molecules based on big

data modeling and analysis. In 2012, Merck supported a QSAR ML challenge to observe the advantages of DL in the drug discovery process in the pharmaceutical industry. DL models showed significant predictivity compared with traditional ML approaches for 15 absorption, distribution, metabolism, excretion, and toxicity (ADMET) data sets of drug candidates .

AI in drug screening :

The process of discovering and developing a drug can take over a decade and costs US$2.8 billion on average. Even then, nine out of ten therapeutic molecules fail Phase II clinical trials and regulatory approval . Algorithms, such as Nearest-Neighbour classifiers, RF, extreme learning machines, SVMs, and deep neural networks (DNNs), are used for VS based on synthesis feasibility and can also predict in vivo activity and toxicity . Several biopharmaceutical companies, such as Bayer, Roche, and Pfizer, have teamed up with IT companies to develop a platform for the discovery of therapies in areas such as immuno-oncology and cardiovascular diseases . The aspects of VS to which AI has been applied are discussed below.

Table 1. Examples of AI tools used in drug discovery

Tools	**Details**	**Website URL**	
DeepChem	MLP model that uses a python- based AI system to find a suitable candidate in drug discovery	https://github.com/deepchem/deepchem	
DeepTox	Software that predicts the toxicity of total of 12 000 drugs	www.bioinf.jku.at/research/DeepTox	
DeepNeuralNetQS AR	Python-based system driven by computational tools that aid detection of the molecular activity of compounds	https://github.com/Merck/DeepNeuralNet-QSAR	

ORGANIC	A molecular generation tool that helps to create molecules with desired properties	https://github.com/aspuru-guzik-group/ORGANIC	
PotentialNet	Uses NNs to predict binding affinity of ligands	https://pubs.acs.org/doi/full/10.1021/acscentsci.8b00507	
Hit Dexter	ML technique to predict molecules that might respond to biochemical assays	http://hitdexter2.zbh. uni-hamburg.de	
DeltaVina	A scoring function for rescoring drug– ligand binding affinity	https://github.com/chengwang88/deltavi na	
Neural graph fingerprint	Helps to predict properties of novel molecules	https://github.com/HIPS/neural- fingerprint	
AlphaFold	Predicts 3D structures of proteins	https://deepmind.com/blog/alphafold	
Chemputer	Helps to report procedure for chemical synthesis in standardized format	https://zenodo.org/record/1481731	[4].

QSAR/QSPR and structure-based modeling with artificial intelligence :

QSAR/QSPR modeling has come a long way since its inception more than 50 years ago
. The impact of these computational models on drug discovery is undeniable, evidenced by the successful prediction of biological activity and pharmacokinetic parameters, viz. absorption, distribution, metabolism, excretion, and toxicity (ADMET) . For ligand-based QSAR/QSPR modeling, the structural features of molecules (*e.g.* as pharmacophore distribution, physicochemical properties, and functional groups) are commonly converted into machine-readable numbers using the so-called molecular descriptors . The spectrum of hand-crafted molecular descriptors is wide, aiming to capture a variety of aspects of the underlying chemical structure. In general, QSAR/QSPR approaches have transitioned from the use of simpler models, such as linear regression and *k*-nearest neighbors, toward more universally applicable machine learning techniques, such as support vector machines (SVM) and gradient boosting methods (GBM) , aiming to address more complex and potentially nonlinear

relationships between the chemical structure and its physicochemical/biological properties, often at the expenses of intepretability [5].

Introduction to Molecular Modeling Methods

Computer experiments play an increasingly significant role in science today. The advent of high-performance computing has enabled virtual experimentation in silico as a tool which allows for interpolation between laboratory experiments and theory. Schulten introduced the term "computational microscope" to describe the role of computational simulations in augmenting experimental research when direct measurements are not possible. He believed that computational biophysics has progressed to the point where it presents a realistic view of intra-cellular components, often at a resolution not attainable through laboratory instruments, reaching atomic or even electronic dimensions. Feynman presciently stated in 1964: "Certainly no subject or field is making more progress on so many fronts at the present moment than biology, and if we were to name the most powerful assumption of all, which leads one on and on in an attempt to understand life, it is that all things are made of atoms, and that everything that living things do can be understood in terms of the jigglings and wigglings of atoms" . Molecular dynamics (MD) is an important computational tool for understanding the physical basis of the structure, the dynamic evolution of the system, and the function of biological macromolecules. Fourteen years later, the first MD simulation of a biological macromolecule, namely, Bovine pancreatic trypsin inhibitor (BPTI), was published . Although the relatively accurate X-ray structure of BPTI was available at the time, its physiological function was unknown[6].

Use of machine learning approaches for novel drug discovery :

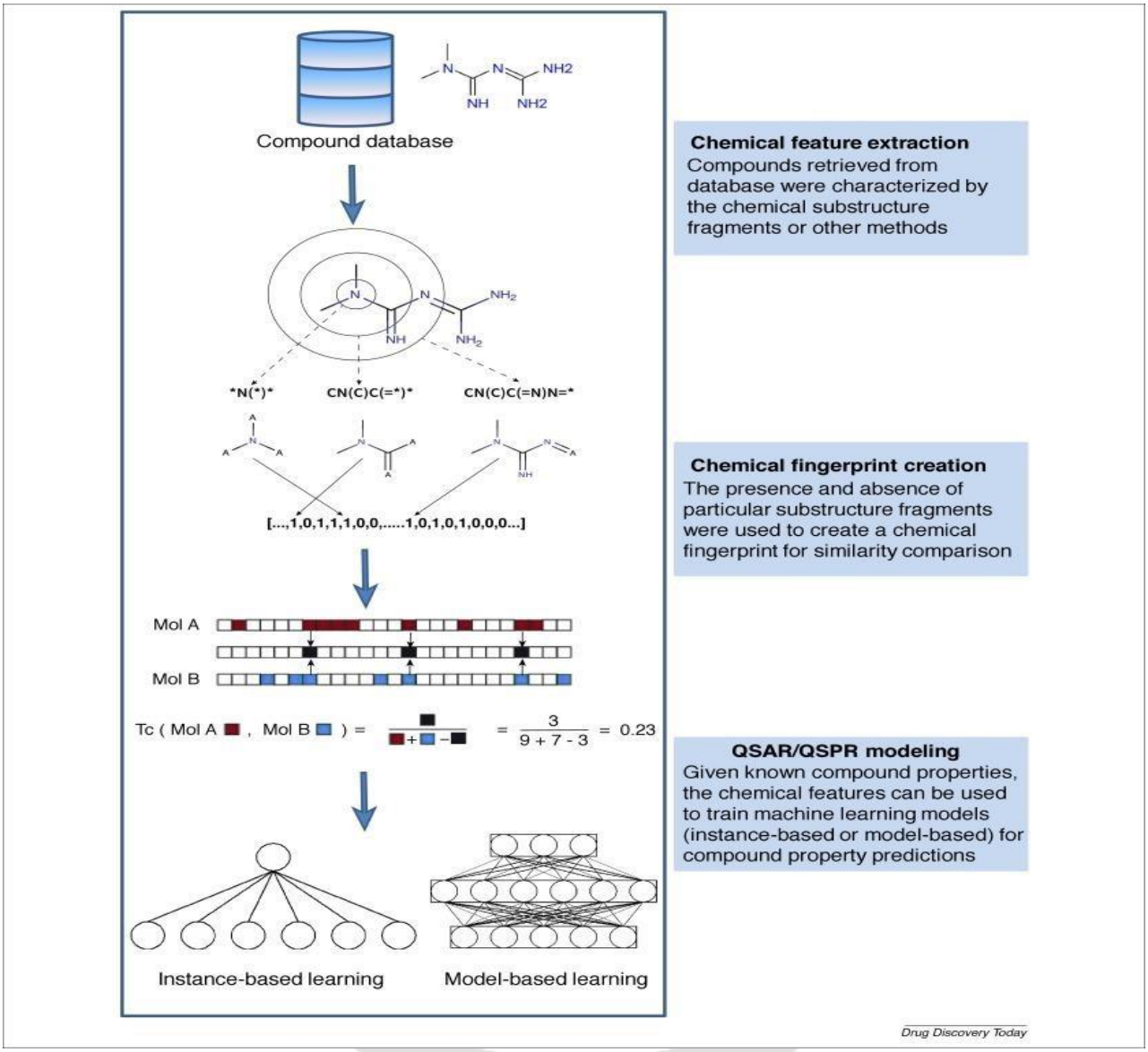
The use of computational tools in the early stages of drug development has increased in recent decades. Machine learning (ML) approaches have been of special interest, since they can be applied in several steps of the drug discovery methodology, such as prediction of target structure, prediction of biological activity of new ligands through model construction, discovery or optimization of hits, and construction of models that predict the pharmacokinetic and toxicological (ADMET) profile of compounds.

Successful cases have been reported in the literature, demonstrating the efficiency of ML techniques combined with traditional approaches to study medicinal chemistry problems. Some ML techniques used in drug design are: support vector machine, random forest, decision trees and artificial neural networks. Currently, an important application of ML techniques is related to the calculation of scoring functions used in docking and virtual screening assays from a consensus, combining traditional and ML techniques in order to improve the prediction of binding sites and docking solutions [7].

Machine learning is currently one of the most important and rapidly evolving topics in computer-aided drug discovery . In contrast to physical models that rely on explicit physical equations like quantum chemistry or molecular dynamics simulations, machine learning approaches use pattern recognition algorithms to discern mathematical relationships between empirical observations of small molecules and extrapolate them to predict chemical, biological and physical properties of novel compounds. Also, in comparison to physical models, machine learning techniques are more efficient and can easily be scaled to big datasets without the need for extensive computational resources. One of the primary application areas for machine learning in drug discovery is helping researchers understand and exploit relationships between chemical structures and their biological activities or SAR . For instance, given a hit compound from a drug screening campaign, we might wish to know how its chemical structure can be optimized to improve its binding affinity, biological responses or physiochemical properties. Fifty years ago, this type of problem could only be addressed through numerous costly, time- consuming, labor-intensive cycles of medicinal chemistry synthesis and analysis.
Today, modern machine learning techniques can be used to model QSAR, or quantitative structure–property relationships (QSPR), and develop artificial intelligence programs that accurately predict *in silico* how chemical modifications might influence biological behavior . Many physiochemical properties of drugs, such as toxicity, metabolism, drug–drug interactions and carcinogenesis, have been effectively modeled by QSAR techniques . Early QSAR models, such as Hansch and Free–Wilson analysis, used simple multivariate regression models to correlate potency (logIC50) with substructure motifs and chemical properties like solubility (logP), hydrophobicity, substituent pattern and electronic factors . Although groundbreaking and successful, these approaches were ultimately limited by unavailability of experimental data and the linearity assumption made in modeling. Therefore, advanced chemoinformatics and machine learning techniques capable of modeling nonlinear datasets, as well as big

data of increasing depth and complexity, are needed[8].

(figure 2).



- Download : Download high-res image (1MB)
- Download : Download full-size image

Figure 2. Computational workflow for chemoinformatics analysis using machine learning. The first step of chemoinformatics analysis is feature extraction, through which the compound is characterized by substructure fragments or other chemical descriptors (first box). The chemical features of the compound are represented by chemical fingerprints and applied for compound similarity comparison based on the presence and absence of shared chemical features. The chemical fingerprint can be used for predicting other chemical and physiochemical properties in QSAR/QSPR analysis using diverse machine learning models including making inference from the training data by comparison (instance-based learning) or from the trained statistical model (model-based learning) (second box).[8]

Low Data Drug Discovery with One-Shot Learning :

The lead optimization step of drug discovery is fundamentally a low-data problem. When biological studies yield evidence that a particular molecule can modulate essential pathways to achieve therapeutic activity, the discovered molecule often fails as a potential drug for a number of reasons including toxicity, low activity, and low solubility. (-3) The central problem of small-molecule based drug-discovery is to optimize the candidate molecule by finding analogue molecules with increased pharmaceutical activity and reduced risks to the patient. Yet, with only a small amount of biological data available on the candidate and related molecules, it is challenging to form accurate predictions for novel compounds[9].

Computational methods in drug discovery :

CADD technologies are powerful tools that can reduce the number of ligands that need to be screened in experimental assays. The most popular complementary approach to HTS is the use of virtual (i.e., in silico) HTS. Computer-aided drug discovery and design not only reduces the costs associated with drug discovery by ensuring that best possible lead compound enters animal studies, but it may also reduce the time it takes for a drug to reach the consumer market. It acts as a "virtual shortcut" in the drug discovery pipeline.

CADD methods can be broadly classified into two groups, namely structure-based (SB) and ligand-based (LB) drug discovery . The CADD method used depends on the availability of target structure information. In order to use SBDD tools, information about target structures needs to be known. Target information is usually obtained experimentally by X-ray crystallography or NMR (nuclear magnetic resonance). When neither is available, computational methods such as homology modeling may be used to predict the three-dimensional structures of targets[10].

The impact of artificial intelligence in medicine on the future role of the physician :

The rise of AI in the era of big data can assist physicians in improving the quality of patient care and provide radiologists with tools for improving the accuracy and efficiency of diagnosis and treatment. AI is well-suited to handle repetitive work processes, managing large amounts of data, and can provide another layer of decision support to mitigate errors. The research firm Frost & Sullivan estimates that AI has the potential to improve patient outcomes by 30% to 40% while reducing treatment costs by up to 50%[11] .

Artificial intelligence with multi-functional machine learning platform development for better healthcare and precision medicine :

Precision medicine is one of the recent and powerful developments in medical care, which has the potential to improve the traditional symptom-driven practice of medicine, allowing earlier interventions using advanced diagnostics and tailoring better and economically personalized treatments. Identifying the best pathway to personalized and population medicine involves the ability to analyze comprehensive patient information together with broader aspects to monitor and distinguish between sick and relatively healthy people, which will lead to a better understanding of biological indicators that can signal shifts in health[12].

Data Science and AI-Based Optimization in Scientific Programming :

Data science and AI-based optimization have also largely been used to solve problems related to scientific programming. Various examples are reported by the literature on task assignment in distributed/parallel systems, knowledge discovery, large-scale data mining, high-performance computing, big data, distributed/parallel search, text analysis/process/classification, and optimization for manufacturing, scheduling, and civil and financial engineering, among others. In this sense, this area provides a wide set of research lines and applications that deserves to be explored[13].

The Stages of Drug Discovery and Development Process :

Drug discovery is a process which aims at identifying a compound therapeutically useful in curing and treating disease. This process involves the identification of candidates, synthesis, characterization, validation, optimization, screening and assays for therapeutic efficacy. Once a compound has shown its significance in these investigations, it will initiate the process of drug development earlier to clinical trials[14].

Computer-Aided Drug Design Methods :

Computational approaches are useful tools to interpret and guide experiments to expedite the antibiotic drug design process. Structure-based drug design (SBDD) and ligand-based drug design (LBDD) are the two general types of computer-aided drug design (CADD) approaches in existence. SBDD methods analyze macromolecular target 3-dimensional structural information, typically of proteins or RNA, to identify key sites and interactions that are important for their respective biological functions. Such information can then be utilized to design antibiotic drugs that can compete with essential interactions involving the target and thus interrupt the biological pathways essential for survival of the microorganism(s)[15].

Modelling three-dimensional protein structures for applications in drug design :

In the absence of experimental 3D structures, protein structure prediction often offers a suitable alternative to facilitate structure-based studies. recent methodical advances in homology modelling, with a focus on those techniques that necessitate consideration of ligand binding. In this context, model quality estimation deserves special attention because the accuracy and reliability of different structure prediction techniques vary considerably, and the quality of a model ultimately determines its usefulness for structure-based drug discovery[16].

It focused on enabling that active participation by helping physicians gain understanding of the core concepts, issues, and trends related to AI (using the common board use of the term which includes Machine Learning, Deep Learning, Augmented Intelligence, and Artificial General Intelligence)[17].

Big data in medicine: The upcoming artificial intelligence:

The new data paradigm in healthcare will be a **"bottom-up" data management strategy** that involves a three-step process after the data is acquired: 1) **data extraction** with various data warehouses providing data; 2) **data transformation** with data configured to a uniform format; and 3) **data loading** with the data entered into an analytical system for final analysis[18].

Advanced machine-learning techniques in drug discovery :

Table 2. Examples of pharmaceutical companies in which ML is central to their business model

Company	**Application**	
AIQ Solutions	Patient heterogeneity	
Atomwise	*De novo* drug	
Benevolent AI	*De novo* drug	
Bioxcel Therapeutics	Clinical trials	
BullFrog AI	Identifying niche patient population	
CytoReason	Target discovery	
DeepCure	*De novo* small molecules	

EVQLV	Biologics discovery	
Genesis Therapeutics	Therapeutic target	
Genome Biologics	Drug repurposing	
Genomenon	Drug discovery and clinical trials	
HealX	Therapeutic target	
InSilico Medicine	Therapeutic target	
InveniAI	Disease indication	
Kintai Therapeutics	*De novo* compounds	
MabSilico	Biomarker discovery	
MIMS	General purpose	
Pepticom	Therapeutic targets	
Recursion	*De novo* drug	
Reveal Biosciences	Pathology	
Socium Inc	Biomarker identification	
Standigm	*De novo* drug	
Thinkcyte	General purpose	
Trials.ai	Designing clinical trials	
TWOXAR	Small-molecule discovery and development [19].	

Artificial intelligence in the pharmaceutical sector :

The current degree of advancement in medical plans and assembling of these items can't address the issues of customized medication. Novel manufacturing assembling arrangements and enabling the adaptable assembling of customized machinery and technology is required in the pharmaceutical industry[20].

References :

1. Abhimanyu S. Ahuja doi :10.7717/peerj.7702
2. Alexandra T. Greenhill MD, CCFP a b, Bethany R. Edmunds PhD c https://doi.org/10.1016/j.tgie.2019.150642
3. ACS Cent. Sci. 2017, 3, 4, 283–293

Publication Date:April 3, 2017 https://doi.org/10.1021/acscentsci.6b00367
4. Maral Aminpour 1 2 3, Carlo Montemagno 4 5 6, Jack A Tuszynski 7 8 9 Affiliations expand PMID: 31052253 PMCID: PMC6539951 DOI: 10.3390/molecules24091693

5. Amol B Deore Jayprabha R
Dhumane Rushikesh Wagh
Vishwakarma Institute Of Information Technology, Pune Rushikesh Sonawane
DOI:10.22270/ajprd.v7i6.616

6. Angélica Nakagawa Lima,Eric Allison Philot,Gustavo Henrique Goulart Trossini,Luis Paulo Barbour Scott,Vinícius Gonçalves Maltarollo &Kathia Maria Honorio

https://doi.org/10.1517/17460441.2016.1146250

7. Anthony C. Chang MD, MBA, MPH, MS
https://doi.org/10.1016/j.ppedcard.2016.08.021
8. Chayna Sarkar 1, Biswadeep Das 2, Vikram Singh Rawat 3, Julie Birdie Wahlang 1, Arvind Nongpiur 4, Iadarilang Tiewsoh 5, Nari M Lyngdoh 6, Debasmita Das 7, Manjunath Bidarolli 2, Hannah Theresa Sony 2

Affiliations expand

PMID: 36768346 PMCID: PMC9916967 DOI: 10.3390/ijms24032026

9. Dnyaneshwar Kalyane 1, Gaurav Sanap 1, Debleena Paul 1, Snehal Shenoy 1, Neelima Anup 1, Suryanarayana Polaka 1, Vishakha Tambe 1, Rakesh K. Tekade 1 2
https://doi.org/10.1016/B978-0-12-814455-8.00003-7
10. H .B. Bode

Beilstein J. Org. Chem. 2016, 12, 2694–2718. https://doi.org/10.3762/bjoc.12.267

11. Isha Rani, Kavita Munjal, Rajeev K. Singla, Rupesh K. Gautam Book Editor(s):Vivek Chavda, Krishnan Anand, Vasso Apostolopoulos
First published: 07 February 2023 https://doi.org/10.1002/9781119865728.ch2

12. 1Juan A. Gómez-Pulido,

2Stéphane Caro,3and José M. Lanza-Gutiérrez4
https://doi.org/10.1155/2019/7154765
13. José Jiménez-Luna 1, Francesca Grisoni 1, Nils Weskamp 2, Gisbert Schneider 1 Affiliations expand
PMID: 33779453 DOI: 10.1080/17460441.2021.1909567

14. Moe Elbadawi 1, Simon Gaisford 1 2, Abdul W. Basit 1 2
https://doi.org/10.1016/j.drudis.2020.12.003
15. Patel Veer 1, Manan Shah 2
https://doi.org/10.1016/j.imed.2021.10.001
16. Paul ‡, Gaurav Sanap ‡, Snehal Shenoy ‡, Dnyaneshwar Kalyane, Kiran Kalia, Rakesh K. Tekade https://doi.org/10.1016/j.drudis.2020.10.010
17. Tobias Schmidt [1], Andreas Bergner [1], Torsten Schwede [2]

DOI: 10.1016/j.drudis.2013.10.027
18. Wenbo Yu 1, Alexander D MacKerell Jr 2 DOI: 10.1007/978-1-4939-6634-9_5
19. Yu-Chen Lo, Stefano E. Rensi, Wen Torng, Russ B. Altman
https://doi.org/10.1016/j.drudis.2018.05.010
20. Zeeshan Ahmed 1 2 3 4, Khalid Mohamed 3, Saman Zeeshan 5, XinQi Dong 1 2 DOI: 10.1093/database/baaa010