IJBPAS, October, Special Issue, 2023, 12(10): 135-161
ISSN: 2277–4998

International Journal of Biology, Pharmacy and Allied Sciences (IJBPAS)
'A Bridge Between Laboratory and Reader'

www.ijbpas.com

THE ERA OF ARTIFICIAL INTELLIGENCE AND BIG DATA: A PATHWAY TO PHARMA 4.0

ASHIK AR, ALEKHYA K*, DAMODHARAN N

Department of Pharmaceutical Quality Assurance, SRM College of Pharmacy, Faculty of Medicine and Health Science, SRM Institute of Science and Technology, Kattankulathur-603203, Chengalpattu District, Tamil Nadu, India

***Corresponding Author: Dr. Kella Alekhya: E Mail: kellaals@srmist.edu.in**

Received 10th May 2023; Revised 6th July 2023; Accepted 6th Sept. 2023; Available online 15th Oct. 2023

https://doi.org/10.31032/IJBPAS/2023/12.10.1018

ABSTRACT

Pharma 4.0 is a new age in the pharmaceutical industry in which Artificial Intelligence (AI) and Big Data (BD) are altering drug discovery, clinical trials, and patient care. The pharmaceutical business is paying close attention to the complete pharmaceutical value chain is being digitally transformed, starting with research and development (R&D) through manufacturing, patient care, sales and marketing, and supply chain management. We concentrated on recent data-driven research developments in domains where AI might effectively reduce costs. Despite pharma's AI expenditures remaining modest, rising funding the cost of AI programs are positive in the big data era.

This article will discuss the following topics:

- an overview of Artificial Intelligence, its subtypes and Big Data;
- the impact of AI and big data in several sectors such as preclinical studies, drug discovery, drug design, clinical trials, pharmacovigilance, chemoinformatics, various diseases, personalized medicines, drug delivery systems and pharma product marketing;
- numerous AI and big data technologies and software implemented in the pharma industry;

- and the challenges and opportunities of implementing AI and big data in the pharmaceutical industry, including regulatory issues, data privacy, and ethical concerns.

This review article will give some understanding of the current situation of Pharma 4.0 and the potential role of AI and big data in revolutionizing the pharmaceutical sector. This also throws a brief outline of current advances in pharmaceutical firms through collaboration with various AI tech companies.

Keywords Artificial intelligence, Big data, Pharma industry, Digitization

1. INTRODUCTION

Since its inception, the pharmaceutical industry has come a long way, and with the advancement of technology, it has entered a new era known as Pharma 4.0. The term itself can be traced back to November 2011, when the German government published an article defining its high-tech strategy, dubbed "Industrie4.0," for 2020. The fourth industrial revolution, or "industry 4.0," combines quickly developing technologies like the Internet of Things (IoT), which includes AI and big data (BD). New thinking will be necessary to achieve Industry 4.0 for pharmaceuticals and overcome the inertia of the present manufacturing infrastructure, operations, and regulation **[1]**.

New terms have emerged as a result of technological advancements in past few years. The utmost popular are big data, Industry 4.0, and artificial intelligence. With the rapid advancement of the young generation of big data and AI technology, an emerging data age-driven "Internet plus AI" is on the horizon **[2]**.

AI imitates or rehabilitates human performance by simulating human intelligence in machine models **[3]**. John McCarthy coined the term in the mid-1950s to explain the integration using science and engineering to build intelligent machines **[4]**. AI operates using machine learning (ML) and deep learning (DL). Machine learning is a subfield of artificial intelligence that deals with the teaching or training of computational models to solve problems or carry out complex tasks using a set of input parameters. Machine learning comes in three categories: reinforcement learning, unsupervised learning, and supervised learning **[5]**.

Classification and regression methods are examples of supervised learning. Unsupervised methods include clustering and feature-finding. ML reinforcement is primarily focused on learning how to complete tasks by comprehending some fundamental principles **[6]**. ML algorithms have proven to be effective at recognizing

and analysing dynamics in massive sizable data, colloquially known as "Big Data". Learning algorithms are trained using big data to improve their recognizing potency and capacity **[7]**.

Artificial neural networks and simulation learning are used in the machine learning subfield known as "deep learning" to build models that can solve complex problems at the same level as the human brain in the real world. Convolutional neural networks (CNNs) and recurrent neural networks (RNNs) are the most common types of artificial neural networks (ANNs) **[2]**. DL techniques are superior to conventional ML techniques in that they can process large amounts of data. As a result, large standardized datasets are urgently needed for DL modelling. When compared to traditional ML methods, DL methods can handle large amounts of data. As a result, there is an urgent need for large standardized datasets for DL modelling **[8]**.

AI also entails the use of computer software applications that analyze, learn, and reveal big data related to pharmaceuticals in order to uncover new drug molecules by incorporating machine learning advancements in a highly unified and mechanized manner.

AI employs a technology-enabled approach that employs a variety of modern networks and tools to simulate the intelligence of people. However, it doesn't raise concerns about fully replacing human existence on Earth physically. AI makes use of tools and software that can decipher, and then instruct the use of input data to produce autonomous results to achieve specific goals **[9]**.

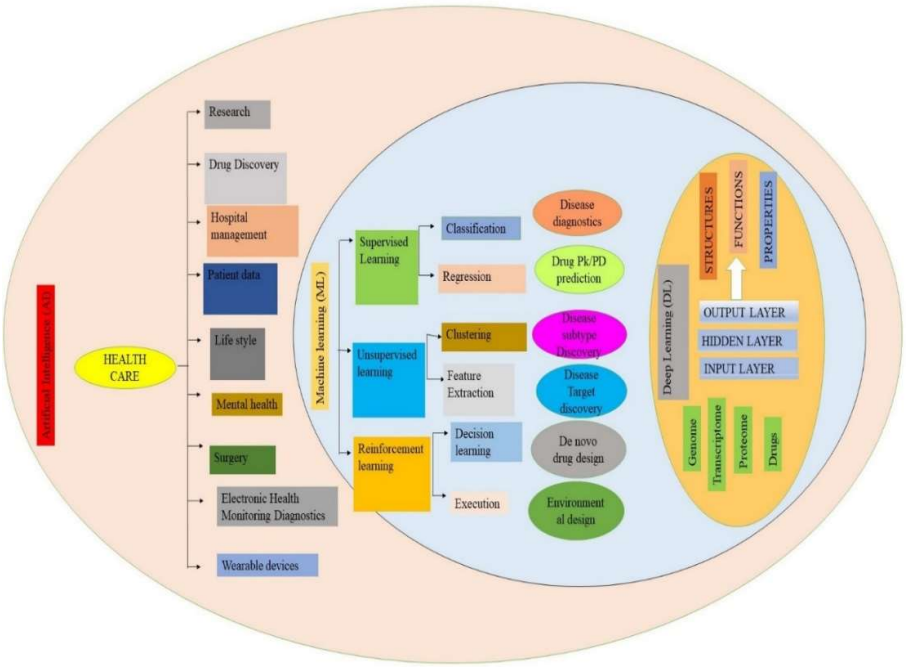

Figure 1: Subfields of AI uses in Healthcare

In recent years, pharmaceutical R&D has become concerned about the potential advantages of utilizing artificial intelligence and all of its related subfields. Over $6 billion is invested in R&D by the top 20 companies, which include Roche, Johnson & Johnson, Merck, Novartis, and Pfizer, the five largest pharmaceutical companies, thus making it the highest investment sector **[10]**. Pharmaceutical R&D still uses stratified adaptive algorithms like neural networks more frequently than deductive or sensing AI because they easily provide computational power to already existing applications. Consequently, cutting-edge computing tools like artificial neural networks have been around for a while and have reached a certain level of maturity in pharmaceutical R&D **[11]**. To fully utilize AI methods, The prevailing pharmaceutical approach needs to change from being expert-driven to being data-driven through collaboration between scientists and AI **[12]**. The increasing number of AI/ML applications demonstrates the potential of big data-driven AI in the pharmaceutical industry; However, the results of ML may not be fully interpretable. As the results produced by ML methods may not be trusted due to their complexity, and they frequently need to be verified by domain experts **[13]**. The application of big data and artificial intelligence (AI) has enabled pharmaceutical companies to process massive amounts of data, predict outcomes, and make informed decisions that are transforming the industry.

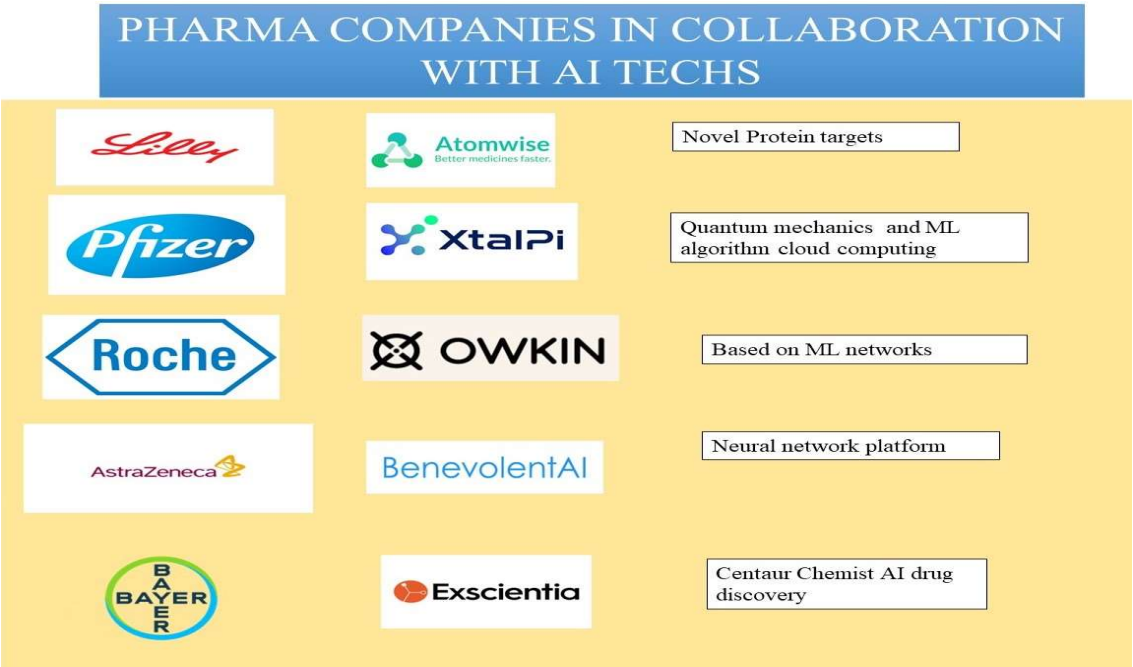

Figure 2: Pharma companies in collaboration with AI Techs

2. ARTIFICIAL INTELLIGENCE IN PHARMA

2.1. Preclinical studies

The modern pharmaceutical industry spends an absurd amount of money due to the ongoing attrition of drug candidates. According to recent statistics, animal toxicity (39 percent), inefficacy (30 percent), and poor pharmacokinetics (39 percent) accounted for 80 percent of the reasons for 11 percent attrition **[14]**. AI techniques have been shown to play a vital role in premarketing drug safety, particularly in the evaluation of toxicity.

Preclinical testing is required to keep dangerous drugs out of clinical trials. This is possible using ML algorithms with interpretable models and low order of complexity. For example, logistic regression & random forest and Deep Learning methods with highly nonlinear models and high orders of complexity. For example, CNNs and RNNs. Thus, it aids in the prediction of toxicity of drugs, modelling patient diversity, informing mechanisms of action, assisting in trial design, drug development, monitoring real-world use, comparative safety analyses, informing precision modelling of patients, assisting in clinical decision-making and regulating actions **[15]**.

According to another study, Histondix employs Genesis 200 which is a multiorgan, quantitative, fully automated, stain-free multiphoton imaging system in conjunction with HistoHepa (a program for automatic image analysis), which is quick, accurate, and error-free. P2K (pattern to knowledge) is a new software assisted by AI, developed by researchers to automatically learn objective information from high-throughput data sets such as Deoxyribonucleic acid, Ribonucleic acid, protein, and antibody interactions **[16]**.

2.2. Drug discovery

A drug takes an average of US$2.8 billion and 12 to 15 years to develop. Traditional methods' low efficacy and high cost have become roadblocks in drug discovery (17). In recent years, computer-assisted drug discovery has successfully used a variety of AI concepts. Deep learning algorithms' capacity to represent intricate nonlinear input-output relationships and carry out pattern recognition and feature extraction from low-level data representations, which are artificial neural networks with multiple processing layers, is largely responsible for this development **[18]**. The stages of the drug research and development process include drug target identification, target validation, hit-to-lead generation, lead optimization, preclinical study, and clinical study **[19]**.

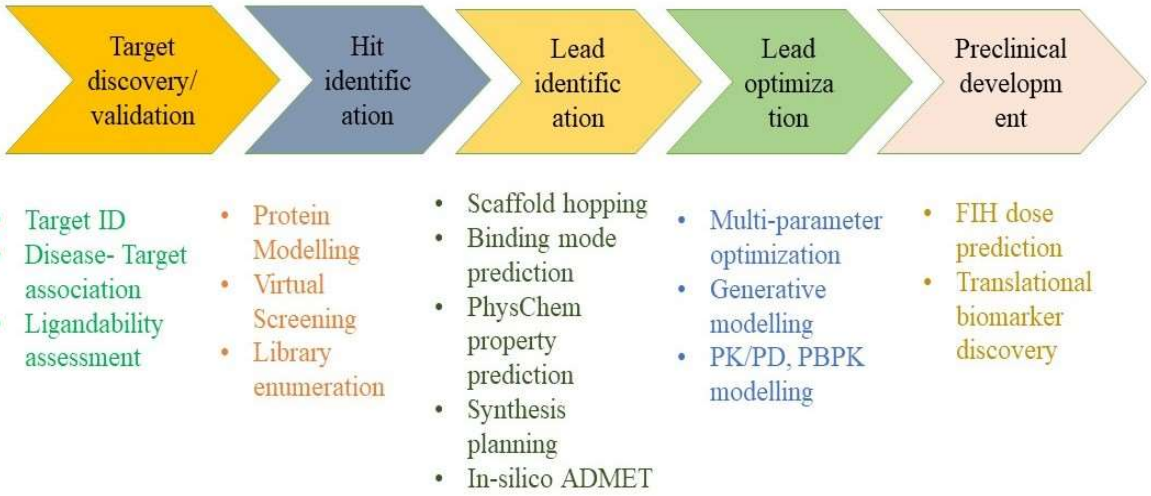

Figure 3: **AI in drug discovery process**

The progression of AI in drug discovery and development was accelerated by the development of cutting-edge chemical descriptors, such as topological descriptors and molecular fingerprints, which substantially expanded the types of descriptors considered from training sets. Drug discovery processes have used AI techniques like De novo drug design, drug-target prediction, and bioavailability prediction. The first neural network proposal in drug discovery was described in the late 1980s. Artificial neural networks (ANN), the first widely used strategy, are based on variable selection. ANNs are a fantastic ML strategy for constructing non-linear interconnections between variables and focusing on biological activities **[20, 21]**.

The adoption and widespread use of virtual screening (VS), quantitative structure-activity relationship (QSAR) modelling and absorption, distribution, metabolism, excretion & toxicity predictive tools have already shown how AI could radically change the discovery and development of drug processes **[22]**.

VS is widely used to aid in rational drug discovery. As AI algorithms become more prevalent in the healthcare and pharmaceutical industries, various models and tools for Ligand-based and Structure-based VS have been developed, including MTiOpenScreen, FlexXScan, and GeauxDock **[23]**. In early drug development, VS can be used to eliminate compounds with the wrong skeletons and as a productive way to locate new hits. As a result, it has become an indispensable means of assisting HTS, abbreviated as high throughout screening; which has issues with high costs and a poor track record of success **[24]**.

Drug design, polypharmacology, drug repurposing, and drug screening are the 4 phases that make up the process of discovering new drugs **[25]**.

Image processing is used to improve the search when creating new drugs or new uses for old ones. These methods are employed to look at biological, chemical & medical databases. Deep semantic analysis of specialized or open data is performed using AI algorithms **[26]**.

Table 1: AI Tools in Drug Discovery

AI TOOLS IN DRUG DISCOVERY	
Deep Chem	**AI technology based on Python is being used to identify a candidate for drug discovery [27]**
Deep Tox	**Predicts the toxicity of over 12000 drugs [28].**
Deep Neural Net QSAR	**a system based on Python that uses computational tools to assist compounds [29].**
Potential Net	**NNs are used to predict ligand binding affinity [30].**
Delta Vina	**A scoring function for a rescoring molecule of drug-ligand binding affinity [31]**
ORGANIC	**A molecular synthesis tool that aids in the creation of molecules with desired properties [9].**

3. CASE STUDY: DL FOR PROTEIN STRUCTURE PREDICTION

The biological mechanism of a protein is a one-dimensional (1D) string of amino acid sequences' ability to encode a three-dimensional (3D) structure.

DeepMind's (Google) **Alpha Fold** is a network of AI that determines a 3D model of a protein based on its amino acid composition. It made protein structure predictions using a DL method based on its sequence. Alpha Fold's central component is a CNN that was trained on structures from the Protein Data Bank to forecast the separations between each pair of residues in a protein sequence. Alpha Fold published structure predictions for Nsp2, Nsp4, Nsp6, SARS-CoV-2 protein membrane and papain-like proteinase which constitute the five SARS-CoV-2 targets in the early 2020s **[32]**.

3.1. Drug design

Big data and artificial intelligence (AI) are now absolutely necessary for drug design due to the huge data accumulation in the abundance of Cheminformatics databases **[33]**. ANNs (artificial neural networks) are heavily used in drug design and molecular modelling. It removes the complication of statistical models used in pharmacokinetic and pharmacodynamic studies, high-throughput virtual screening, and quantitative structure-activity relationships **[3]**.

AI drug design algorithms use two molecular representations: a molecular fingerprint and a simplified molecular input line-entry system string **[34]**. Any genuine AI system for de novo molecular design must possess the basic intelligence to produce novel drug like, and chemically rational molecular representations on its

own. According to Olivecrona and AstraZeneca's recent research, simplified molecular input line entry system representations from the CHEMBL database were used to train an RNN. To groove the pre-trained RNNs and produce new sequences that abided by the conditional probability distributions discovered from the training set, policy-based reinforcement learning was used. The network generated 94 percent of the sequences, which matched to reliable molecular structures, 90% of which were new **[35]**.

The approaches to designing drugs were centred on databases that were built using different ML algorithms. Numerous ML and DL tools for forward synthetic prediction have been developed and implemented to influence how chemical reactions turn out.

To make the prediction job, these tools typically use a combination of statistical learning methods, including the classification of reaction types and the automatic location of the reaction center. The majority of the tools used by present DL methods for retrosynthetic analysis are similar, and their main differences lie in how molecules are represented **[17]**. For relatively small data sets, multiple linear regression and partial least squares are the two straightforward machine learning algorithms that are often used to design a model **[22]**.

To reduce time and manual interference, machine learning techniques are used in the majority of drug design processes. The ideal example is QSAR, where huge dataset training and data collection are viewed as rate-limiting steps when determining ligand-based virtual screening protocols **[36]**. Several publications have appeared in the last few decades that use a variety of ML techniques to model QSAR. Because it is simple to parallelize and produces excellent predictions with few parameters, the random forest algorithm has been a wise substitute. DNNs are currently on the rise. One of the earliest models using feed-forward NNs was QSAR. CNNs, on the other hand, take advantage of the data's hidden structures to produce better results, overcoming the drawbacks of unbalanced datasets. Moreover, Karpov's work as a trailblazer in the QSAR task application of a transformer-based model is noteworthy **[37]**.

A QSAR model is usually developed in four steps:

- gathering a training set of data (namely, chemicals with physical and/or biological characteristics that can be procured observationally),
- embedding chemicals that contain molecular entities (namely, the characteristics of each molecule),
- using mathematical algorithms to train the model to forecast chemical

properties based on their molecular descriptors (ranging from multiple linear regression which is simple to slashing machine learning algorithms),

- using a validation dataset to assess model performance **[5]**.

AI has become an integral feature of QSAR research since it can efficaciously establish the relationship between chemical structure and biological activity that is represented by a resilient model. As a consequence, conventional ML methods including RFBoosting, GP, Cubist, SVM, KNN, and DL been widely employed in building QSAR models **[24]**.

4. Clinical trials

The ability of AI to predict drug properties may also eliminate the need for clinical trials and human study subjects, which would be advantageous from both a financial and ethical perspective **[38]**. AI has the potential to influence practically every aspect of clinical trials, including patient selection, drug adherence monitoring, sharing and use of medical records, analysis, operational processes, and layout **[10]**. The difficulty with clinical prediction is that it is currently impossible to foresee which drug candidates will behave safely and effectively in humans before investing in expensive clinical trials. Sadly, animal testing seems to be an utmost poor predictor of clinical efficacy and safety.

The updated FDA Modernization Act was enacted just before December 31, 2022; repealing an eighty-four-year-old act that required sole dependence on animal studies and substituting it with more contemporary, promising strategies, earning the admiration of animal welfare activists everywhere **[39]**. A further challenge in clinical trials is the large sample size required to utilize DL techniques unequivocally for predictive reactions in therapy. In clinical trials, the DL necessitates a large sample size. It is occasionally possible to integrate data in clinical trials, but this can make it difficult to interpret the results if there is favouritism **[36]**.

Figure 4: Predicting safety using AI in Clinical Trials

Clinical trial datasets on patient characteristics and design can be utilized to inform ML models that predict regulatory approval and indicate the likelihood that phase transitions will be successful. AI technologies are remarkably useful in clinical trial patient selection and large-scale analytics, access equity, and biomarker improvement to support trial-matching search engines.

In a Phase II trial involving individuals with schizophrenia, it was reported that the AI software known as 'AiCure' was developed as a mobile application to estimate medication adherence. AiCure accelerated adherence by 25% in comparison to conventional "modified directly observed therapy." The use of AI predictive modeling to select patients would increase the success rates of clinical trials **[25, 40]**. Using Exscientia's AI-assisted platform, Sumitomo Dainippon Pharmaceutical company in Japan recently developed a drug molecule for obsessive-compulsive disorder in less than a year. The first AI-made molecule intended for human trials is this one **[16]**.

5. Pharmacovigilance

The science and practices involved in the identification, evaluation, cognition, and mitigation of negative effects are referred to as pharmacovigilance. The tasks include DDI (drug-drug interaction) extraction from biomedical literature, DDI classification, and DDI, a condition in which one drug affects the activity of another drug. Text mining uses ML techniques such as feature-based (FB) and kernel-based (KB) methods, which are adapted to biomedical literature and clinical reports for DDI tasks **[41]**.

After trying to predict ADRs for drugs, AI stretches DL applications to the further stage for other forecasting applications, including precision medicine and drug repurposing **[42]**.

6. Chemoinformatics

Virtual screening has long been a part of chemoinformatics, in which computational tools are used to sift through massive databases in search of fresh leads with a high chance that the target protein will have a high binding affinity.

VS can broadly be classified into structured-based VS and ligand-based VS. Structure-based VS is used when there is sufficient information on the three-dimensional (3D) structure, whereas ligand-based VS is used when there is insufficient 3D information. Machine learning approaches, particularly for ligand-based virtual screening, have been integrated into virtual screening **[4]**.

7. Various Diseases

In recent years, artificial intelligence has the ability to contribute to the advancement of the healthcare regime by making it more efficient in identifying illnesses like cancer, infections, and other diseases, as well as the creation of robots that can perform surgeries and precise diagnosis **[2]**.

In cardiovascular disease, ML techniques use blood tests and MRI scans to assess patients' chances of survival rate to improve diagnostic efficiency and optimize the treatment process **[41]**. Using normal sinus rhythm electrocardiographs, AI algorithms can predict the possibility of stroke and atrial fibrillation. Neurologists can identify patients with acute ischemic stroke caused by large vessel blockage as well as real-time candidates for revascularization with the help of computed tomography neuroimaging and MRI scans **[43]**.

In the case of diabetes, apps such as Suguard use AI to create a system of self-management for a patient on what to eat or drink and how much time they should spend exercising, as well as analysis of food calorific value in real-time **[44]**. Random forest models were developed to predict chronic kidney disease (CKD) progression by utilizing EMR data in conjunction with circulating biomarkers like plasma tumour necrosis factors and kidney injury molecule-1. ANN also has been developed to predict the progression of renal failure in CKD patients. The model could predict the estimated glomerular filtration rate (eGFR) with greater than 95% accuracy in 6, 12, and 18-month intervals **[45]**.

AI techniques supply novel insights into cancer biophysics and enable cancer pathway prediction. An artificial intelligence model was used to determine the risk of lymph node metastasis in colorectal cancer patients and determine the requirement for additional surgery following endoscopic resection of the tumour. Unnecessary surgical procedures can be avoided in this regard **[46]**.

AI techniques combined with demographic and clinical information can forecast the course of chronic kidney disease. Using EMR and circulating biomarkers, random forest(RF) models calculate a strong

predictive risk score and the likelihood of developing proteinuria **[41]**. The use of AI in respiratory diseases, combined with the availability of X-ray images and electronic health records, is chiefly used in the early detection and treatment of infections of the lower respiratory tract. Based on the patterns on chest X-rays and the cause, AI could classify viral and bacterial infections and detect pneumonia **[47]**.

8. Personalized Medicines

Uncovering the mechanism of action of drug-target-disease-therapy preclinical and clinical interventions and determining whether a chosen intervention is effective in both a preclinical and clinical scenario has been the main goals of the use of AI-based integrated modelling in the development of personalized medicines for the treatment of people with known diseases. As a result, individual evaluation, prognosis, and therapeutic interventions are the primary focus of AI-integrated modelling platforms used to advance personalized medicine **[48]**. Personalized medicine is now utilizing DL tools to hereditarily customise a person's gene to anticipate future diseases that an individual may contract. Particularly in tackling a vanishing gradient problem, the effectiveness of limited DL in acknowledging multi-labelled learning with inadequate markings was demonstrated **[42]**.

UCSF Medical Center employs robotic technology to prepare and track pharmaceuticals to improve patient safety. In terms of size and ability to administer precise drugs, robots have proven to be significantly superior to humans. The automated system's ability to create sterile chemotherapy preparations and fill intravascular syringes with the appropriate drugs expands its capabilities **[49]**.

9. Drug delivery routes

Integrated circuits, detectors, battery packs, and safe data backups are the main components of AI-based nanorobots, which are often sustained by computational technologies like AI. These are trained to detect collisions, locate targets, detach & attach and then excrete from the body. AI tools like NNs, fuzzy logic, and integrators that control automation is required for drug release. Implants with microchips are used for both coded release and locating the implant in the body **[27]**.

With the aid of software like LAMPS and GROMACS 4, surface chemistry and the internalization of nanoparticles into cells can be investigated **[50]**.

10. Pharma product marketing

Huge statistics data sets, including costs associated with product development, market potential, product cost, production cost, and the prices of competitors' products, are assessed by AI systems such as Incompetitor, which then develops

algorithms for predicting product prices. Pharmaceutical firms are utilizing AI to lower the costs and failure risks related to VS.

The market for AI is expected to reach $5 billion by 2024, up from $200 million in 2015 and $700 million in 2018. AI is anticipated to revolutionize the healthcare and pharmaceutical industries, with a 40 percent projected growth from 2017- 24.

Numerous pharma companies have invested in AI and are continuing to do so. They also work with AI firms to create vital healthcare tools **[27]**. Referring to IQVIA's President, "Machine learning and artificial intelligence enable global life science sales, marketing, and branding teams to develop more profitable and actionable commercialization strategies based on AI insights."

AI/ML also allows healthcare organizations to delve deeper into more detailed layers of HCP, patient, and payer information in order to uncover previous untapped perspectives, predict potential courses of action, and facilitate quick and efficient decision-making **[51]**. AI will be able to deliver nicer process stakeholder-aligned information for the customer, allowing for more targeted information dissemination.

Pharma companies can use AI to test and create innovative marketing plans that assure high revenue and brand awareness **[52]**.

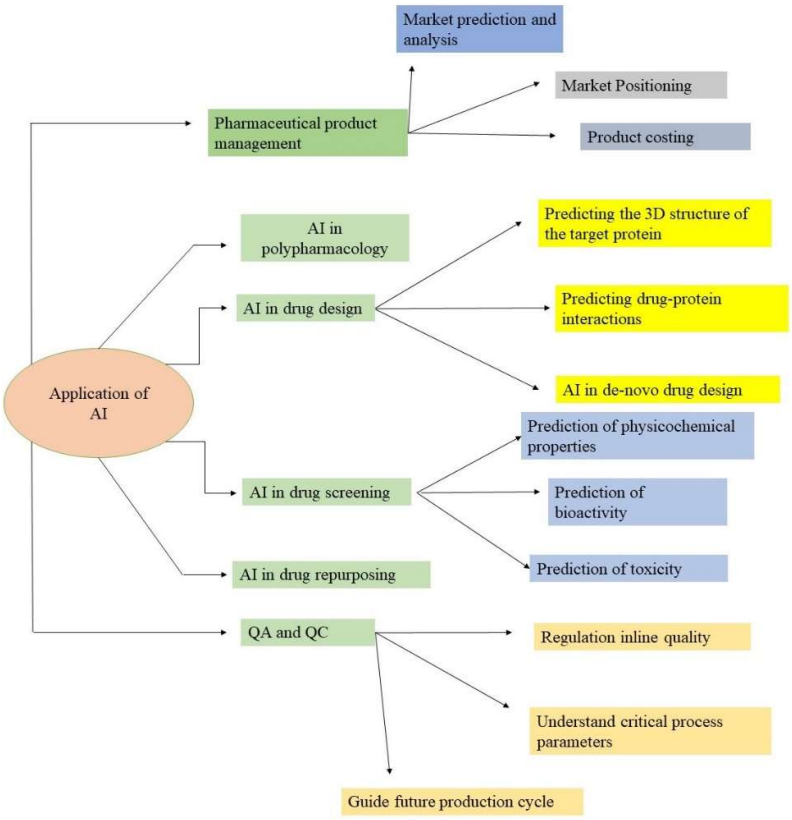

Figure 5: Application of AI in the Pharmaceutical field

11. BIG DATA IN PHARMA

Big Data (BD) is data that necessitates infrastructure or technical investment to produce valuable insights. Understanding the depth of Big Data begins with understanding the characteristics of a given dataset that are defined by these four attributes; Volume, Variety, Velocity, and Veracity are commonly referred to as the "Four Vs". Value, the 5th "V," is frequently mentioned and is arguably the most crucial attribute **[13]**. The majority of the big data surge is unstructured data, which is difficult for traditional databases to analyse. As a result, the predictive power of big data has recently been investigated in fields such as public health, science, and medicine **[53]**. Massive datasets are routinely acquired, stored, and analysed in the era of big data to inform biomedical discoveries & validate hypotheses **[54]**. Big data is becoming more accessible in all areas of manufacturing and operations. Intellectual property analytics (IPA) is a branch of data science that examines vast amounts of data pertaining to intellectual property to identify relations, statistics, and designs that can be used in decision-making. It is an interdisciplinary strategy that employs computer programming, mathematics, statistics, and operations research to extract useful information to help in facilitating business judgments using data **[55]**.

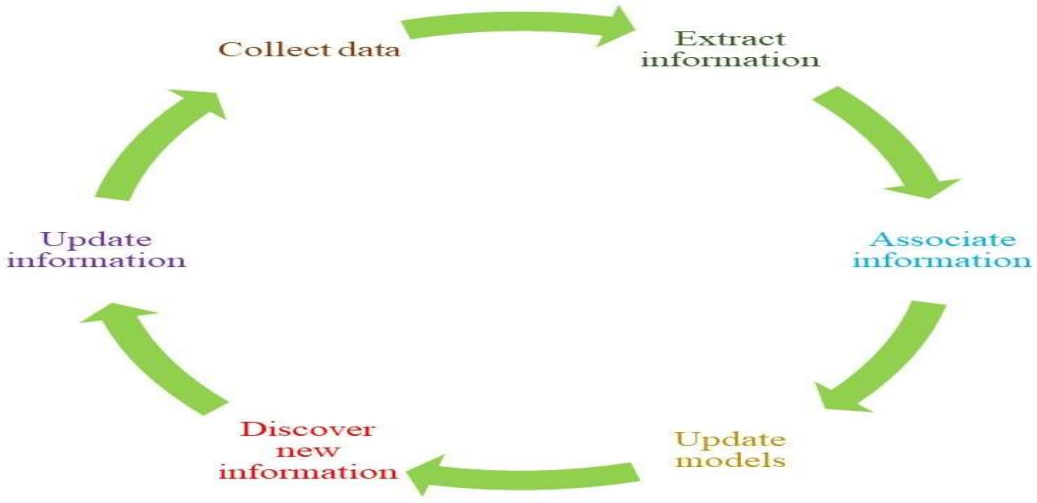

Figure 6: Process in Big data

To handle big data in healthcare-related applications, most applications for healthcare services use hybrid models, which combine cloud technologies and other services. As a result, systematic innovations toward Pharma 4.0 applications are produced. Track-and-trace analytics in healthcare and pharmacy, as well as useful

supply chain analytics, require robust database systems and in-process memory. This idea necessitates a change in computing infrastructures to address the need for data storage and the strong server processing needed for the secure analyzation of huge data volumes **[56]**.

The term "Big Data Analytics" (BDA) in the context of healthcare refers to techniques for analyzing significant amounts of digitized patient health and well-being data. Due to the enormous diversity of this data, conventional software or hardware cannot be used to quantify it. By providing researchers with access to a sizable sample of data, big-data clinical trials will augment and enhance randomized controlled trials in the field of clinical research. Thus, researchers can plan potential clinical trials relying on hypotheses produced by data analysis, analyze the impact of clinical trials, and identify the risks and side effects associated with products before they are released publicly for use due to the advances in data analytics and machine learning techniques.

- BD is a relatively new scenario that describes the procurement of large-scale datasets and their integration using advanced analytics to enable new information or understanding of these data. Following the acquisition, big data may be used to increase the production of pharmaceuticals through a 3-step data management procedure that consists of the subsequent steps:
- extraction and collection of a sizable amount of dispersed and diverse data;
- Organizing data to maintain consistent formatting;
- Last but not least, data analysis using various analytical platforms to provide a final output, the interpretation of which may influence decisions regarding the development of certain substances or medications or the employment of particular techniques **[38]**.

The most popular tool for data analysis was clinical data (70 percent). Hadoop, which was released in the early 2010s by Apache, is primarily comprised of the Hadoop Distributed File System (HDFS) which is a technique that breaks down big data sets into smaller ones and distributes them across multiple servers & MapReduce which is a computational approach that uses parallel processing through two execution sequences **[57, 58]**. Big data analytics combined with ML can be used to integrate heterogeneous data sources **[36]**. Today, clinicaltrials.gov, the largest main trial registry, which keeps data from over 255,000 trials, offers a

special chance to learn about the trial design and competitive environment **[59]**.

Presently, most portals for data storage, such as PubChem gather information from various sources, identifying the types of information. The incongruity in data veracity and quality reflects the level of uncertainty that is there in data from several sources and necessitates innovative data management and curation technologies. The data variety is always accompanied by the features vocabulary and venue. Different formats, scripts, or phrases can be used to describe data from multiple sources. The data vocabulary, that is terminology must be normalized when collecting data from the platform, which is the actual data location **[60]**.

The aggregated big data used for the discovery of drugs can be divided into a variety of categories or databases, such as those that contain a catalogue of chemical compounds like PubChem and ChEMBL, a collection of drugs and drugs-like substances such as Drugbank and e-Drug3D, a collection of drug targets that includes genomic and proteomic data like Binding DB and Supertarget, databases with a list of assay screening, metabolism, and efficacy studies like HMDB and TTD (The (Therapeutic Target Database). Some of the databases used for drug research and discovery are LinkedOmics, STRING, and DisGeNET. Furthermore, there are now so many ML approaches available to help with extracting vital features, connections, and structures contained in these enormous biological data sets, that the rise of AI has made big data analytics considerably simpler **[14, 23, 33]**.

An abundance of biological data is being generated daily as a result of the development of technologies like microarray, RNA sequencing, and high-throughput screening, ushering in the BD era for contemporary drug research. Every day, the clinical and operational information systems used by the healthcare sector, such as electronic health records and laboratory information library systems, manage a substantial volume of data **[23]**.

Over the last decade, quick advancements in high-throughput screening methods and pertinent data-sharing mechanisms have propelled current drug development into the era of big data **[61]**. Big Data-enabled drug discovery faces significant challenges if it is to truly transform how the process of developing new drugs and validating targets is conducted. How Big Data is used to provide answers to these questions is described in the Open PHACTS IMI project's approach, in which massive amounts of data are incorporated to methodically answer crucial issues in the early stages of drug discovery **[59]**.

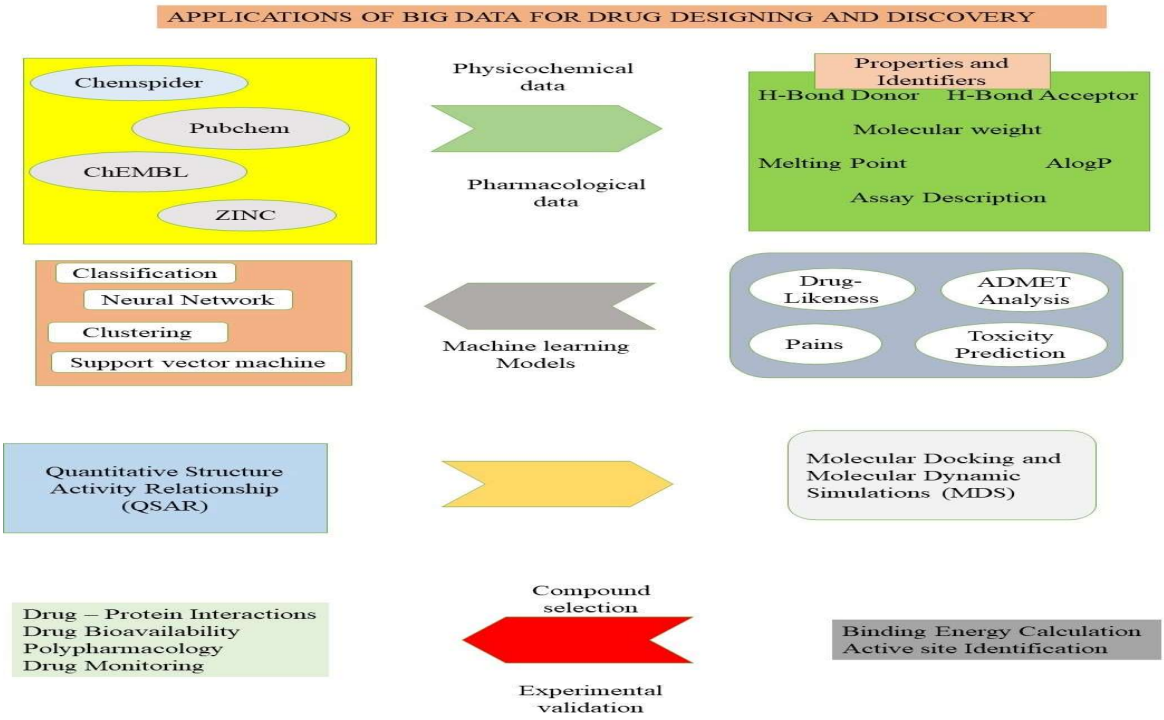

Figure 7: Applications of bigdata for drug designing and discovery

The diversity of multidimensional data sets needed to avoid obstacles on the path from the start to the end of a drug in both a personalized and public context that has increased as a result of advances in multi-omics and BD generated from high-throughput pharmacokinetic & pharmacodynamic analysis of preclinical models and clinical trials. These sets of data are critical for characterizing the impacts on signaling pathways based on molecular networks that support the operations of biological systems and flaws that cause disease pathology and possible customized medicine remedies **[48]**. The century of huge volumes of drug and drug candidate information as a result of technological advancement has transitioned modern drug discovery approaches towards the big data era. The constant growth of big data necessitates more to analyze the generated complicated data, processing power and cutting-edge computational methods.

Big data will undoubtedly enable the identification of system of disease-related signals from genes and epigenetics. There is an appropriate design for integrating data sources from diverse disciplines using Open PHACTS. This contains suggestions for appropriate license models as well as thorough guidelines on how to get data ready for integration **[62]**.

Big data is concerned with massive or intricate data sets that are analysed systematically utilizing data analysis methods such as machine learning and data

mining from various sources. For instance, big data has sparked widespread interest in academia for its potential to assist both patients and medical personnel by enhancing the precision of Making decisions, making a diagnosis, and predicting disease **[63]**. The DNN models' successful outcome demonstrates the benefits of using DL techniques for modeling massive data sets and choosing pertinent features **[21]**.

Data mining, a type of big data analysis, has a cornerstone in monitoring and managing the global broadcast of pandemic diseases like COVID-19. Big data analysis can help tackle issues like trajectory modelling and the discovery of COVID-19 spatiotemporal patterns. Big data and AI can help with monitoring liberalization, epidemic alerts and management, trend analysis and evaluation, identifying the source of an infection, allocation of a resource, prevention and its diagnosis, and vaccine & drug development **[64]**.

Through electronic health records (EHR) or health monitoring, Big Tech companies have already begun collaborating with pharma companies in the areas of data analytics, diagnostics, patient monitoring, and real-world proof. To illustrate this in real-time, A project called "Baseline" has been started, where "Verily" is working with Pfizer, Otsuka, Novartis, and Sanofi, utilizing its technology to boost patient and clinician involvement, mitigate study time, and produce more valuable insights to enhance clinical research **[65]**.

12. CHALLENGES

For starters, the AI model mechanism is ambiguous. AI techniques are commonly known as "black boxes." (How AI platforms arrive at their conclusions)". There are limited methods for explaining demonstrations and a scarcity of plausible explanations for relevant biological processes. Second, there is excessive fit and the requirement for enormous sets of data. These belong to the 'Narrow AI' category.

DL approaches are ineffective because they can't produce a fair estimate of generalization from very little data. Hence, they may be less practical than some traditional shallow ML methods **[8]**. The necessity for a significant volume of high-quality training data is one of the most challenging aspects of integrating AI into drug research.

Another challenge in drug discovery is the transparency and interpretability of AI and big data, which implies that these models' characteristics, operations, and weights are beyond the comprehension of a human user. Because making a decision is ambiguous in various classification models **[66]**. To explain the findings, many pathways in drug development must be recognized. As a result, finding new drug targets is made easier, and trust in interpretability must

improve due to the various combined features.

Furthermore, training the neural network model demands extensive parameter tweaking. There is eventually the issue of cost calculation. Even though AI models require low computing resources to train, the training process, particularly for DL models with more hidden units, is computationally intensive and takes a lot of time. To process some large data sets, GPU is also needed, yielding a result of computational expenses to be quite expensive **[24]**. When attempts are made to adopt requisite ethical standards for artificial intelligence, medical ethical issues arise **[7]**.

Using Big Data ideas is a difficult task for pharmaceutical companies because of the constant stress and commercial needs that are created by the adoption of standards, research requirements, and legal requirements. These legal and standard requirements are linked to drug control and the safety of human healthcare, which necessitates ongoing and in-depth data analysis **[56]**. Large-scale current biological data that is high-dimensional, sophisticated, and diverse is too difficult to analyze using human-dominated classical models in the big data age.

The lack of scores for the target disease association need validation or benchmarking., as well as target draggability information, is a major limitation of Big Data in data curation. The quantity of substantiating data sources and arguments has no direct relationship with the target modification's effectiveness in drug discovery. As a result, the scores must be validated through experiments or benchmark studies. Most databases merely display the amount of data sources' supporting evidence or the supporting evidence itself.**[14]**.

Other barriers to AI and big data include:

- Regulatory issues such as a lack of clear guidelines, data validation, bias, and fairness.
- Patient consent, data breaches, and data ownership are all examples of data privacy.
- Ethical concerns, such as equity and autonomy.

13. CONCLUSION

Artificial intelligence and Big Data can contribute to the development of new technologies and software that will aid in the development of pharmaceutical products and health management strategies **[52]**. Medical devices based on artificial intelligence and machine learning (AI/ML) strive to achieve patient support by identifying fresh information derived from large data produced by individual patients and the culmination of numerous sufferers' experiences **[67]**. The incorporation of AI to find new drugs has large potential to

produce hopeful outcomes in the research and development of novel cures for fatal diseases. Additionally, it will cut research and development expenses while raising the success rates of drug discovery **[16].**

Throughout all areas of the pharma industry, the discovery of drugs has been the most heavily concentrated in terms of incorporating AI and big data through collaboration with large tech companies. In today's big data era, strategies from ML and AI have been used to analyze data produced at various stages of drug research and development; confirming the value of BD by lowering drug attrition. These hurdles posed by big data's multiple Vs feature necessitate the development of appropriate computational approaches and algorithms **[60]**.

AI and machine learning aid in the discovery of drugs by utilizing in evaluating synthetic routes, designing compounds in silico, and simulating absorption, distribution, metabolism &elimination and binding characteristics. Recent developments include the use of absorption, distribution, metabolism & elimination and ligand binding models, as well as newer machine learning techniques, in the construction of tools for structure generation and synthetic assessment **[22]**.

With the advent of AI and the elevated use of many techniques, including deep learning and machine learning, these have been used to analyze intellectual property data **[55]**. A tailored drug with the desired dose, delivery parameters, and other necessary components can be manufactured depending on the needs of each patient by integrating AI into the manufacturing of pharmaceutical products **[68]**. The importance of automation will be further highlighted by using the most recent AI-based technologies, which will also improve the quality and general safety of the manufacturing process and better utilize the available resources while remaining cost-effective. This will reduce the time it takes for products to reach the market.

Since BD has been around for a while and the pharmaceutical sector has traditionally used complex data sets, Only recently have all of the Big Data analytics tools become available are gradually coming together. The most significant progress has been made in the field of research on outcomes and health economics, as demonstrated by Sanofi & Pfizer in multiple situations **[69]**. When offering BD solutions, a non-competitive integrated strategy would result in considerable cost reductions, especially with regard to publicly available data **[62]**.

AI and big data have numerous applications in the pharmaceutical industry. AI algorithms, for example, can be used to evaluate massive number of data from clinical trials and drug development, assisting researchers in identifying new drug targets and revamping the drug discovery

process. Furthermore, AI-powered systems can assist pharmaceutical companies in personalizing medicine and developing precision therapies based on genetic make-up, medical history, and other characteristics specific to patients. Furthermore, big data analytics can provide insightful information about patient behaviour, disease trends, and treatment outcomes. Pharmaceutical companies can identify areas of unmet medical need and develop innovative solutions to improve patient outcomes by analysing this data. Besides that, big data analytics can also aid in the maximization of clinical trial design, the reduction of costs, and the acceleration of drug development timelines.

AI systems must follow ethical principles and consider if software programs are medical devices that need to follow formal regulatory processes and undergo initial trials. Additionally, before incorporating AI solutions at the point of care, regulations and policies for the delivery of outputs to patients and clinicians must be established **[45]**.

Nevertheless, the convergence of AI and big data in the pharmaceutical industry is fraught with difficulties. Large amounts of data necessitate complex algorithms and powerful computing infrastructure, which can be expensive and difficult to implement. Furthermore, the application of AI and big data raises ethical concerns about data privacy, security, and bias. Pharmaceutical companies can address these challenges and ensure the safety, security, and ethics of their AI and big data applications by collaborating closely with regulatory agencies, patient advocacy groups, and ethical experts.

AI recognition technologies are more in requirement in the pharmaceutical industry at the R&D stage, whereas AI analytics technologies are more in consumption at the production and sales stages **[26]**. Moreover, if the hype surrounding digital health could be turned into optimism, people would be more open and trusting of the pharmaceutical sector, which would change how future therapies are created and demonstrate the industry's importance to society **[70]**.

From this review, we can conclude that the impact of AI and BD in the pharma sector is undeniable one. Right from animal testing to the marketing of drugs, these aid in reducing time, costs and failure of trials. Developing a new drug molecule will no longer be a hectic process, thus giving a ray of hope to treat all deadly diseases. Optimal therapeutic dose with lesser defects of drug manufacturing which is being prescribed based on individual person lifestyle is seen to be the biggest achievement of these technologies. All these indicate that the industry is currently shifting towards the concept of Pharma 4.0. Despite all these

obstacles, the potential benefits of Pharma 4.0 are enormous, and the pharmaceutical industry is heavily investing in AI and big data technologies. In the big data era, AI technology is anticipated to soon reach all sections of the pharmaceutical industry due to the continuing gathering of medical data and the development of increasingly sophisticated AI algorithms.

REFERENCES

[1] Arden NS, Fisher AC, Tyner K, Yu LX, Lee SL, Kopcha M. Industry 4.0 for pharmaceutical manufacturing: Preparing for the smart factories of the future. Vol. 602, International Journal of Pharmaceutics. Elsevier B.V.; 2021.

[2] Aggarwal K, Mijwil MM, Sonia, Al-Mistarehi AH, Alomari S, Gök M, *et al*. Has the Future Started? The Current Growth of Artificial Intelligence, Machine Learning, and Deep Learning. Iraqi Journal for Computer Science and Mathematics. 2022;3(1):115–123.

[3] Selvaraj C, Chandra I, Singh SK. Artificial intelligence and machine learning approaches for drug design: challenges and opportunities for the pharmaceutical industries. Vol. 26, Molecular Diversity. Springer Science and Business Media Deutschland GmbH; 2022. p. 1893–1913.

[4] Rashid MBMA. Artificial Intelligence Effecting a Paradigm Shift in Drug Development. Vol. 26, SLAS Technology. SAGE Publications Inc.; 2021. p. 3–15.

[5] Lin Z, Chou WC. Machine Learning and Artificial Intelligence in Toxicological Sciences. Vol. 189, Toxicological sciences : an official journal of the Society of Toxicology. NLM (Medline); 2022. p. 7–19.

[6] Toh TS, Dondelinger F, Wang D. Looking beyond the hype: Applied AI and machine learning in translational medicine. EBioMedicine. 2019 Sep 1;47:607–615.

[7] Keskinbora KH. Medical ethics considerations on artificial intelligence. Vol. 64, Journal of Clinical Neuroscience. Churchill Livingstone; 2019. p. 277–282.

[8] Jing Y, Bian Y, Hu Z, Wang L, Xie XQS. Deep Learning for Drug Design: an Artificial Intelligence Paradigm for Drug Discovery in the Big Data Era. Vol. 20, AAPS Journal. Springer New York LLC; 2018.

[9] Sarkar C, Das B, Rawat VS, Wahlang JB, Nongpiur A, Tiewsoh I, *et al*. Artificial Intelligence and Machine Learning Technology Driven Modern Drug Discovery and Development. Vol. 24, International Journal of Molecular Sciences. MDPI; 2023.

[10] Henstock P. Archives of Pharmacology and Therapeutics Commentary https://www.scientificarchives.com/journal/archives-of-pharmacology-and-therapeutics Artificial Intelligence in Pharma: Positive Trends but More Investment Needed to Drive a Transformation [Internet]. Vol. 2, Arch Pharmacol Ther. 2020. Available from: https://www.scientificarchives.com/journal/archives-of-pharmacology-and-therapeutics

[11] Colombo S. Applications of artificial intelligence in drug delivery and pharmaceutical development. In: Artificial Intelligence in Healthcare. Elsevier; 2020. p. 85–116.

[12] Henstock P V. Science and Society Artificial Intelligence for Pharma: Time for Internal Investment [Internet]. Vol. 40, Trends in Pharmacological Sciences. 2019.

[13] Cremin CJ, Dash S, Huang X. Big data: Historic advances and emerging trends in biomedical research. Curr Res Biotechnol. 2022 Jan 1;4:138–151.

[14] Kim H, Kim E, Lee I, Bae B, Park M, Nam H. Artificial Intelligence in Drug Discovery: A Comprehensive Review of Data-driven and Machine Learning Approaches. Vol. 25, Biotechnology and Bioprocess Engineering. Korean Society for Biotechnology and Bioengineering; 2020. p. 895–930.

[15] Basile AO, Yahi A, Tatonetti NP. Artificial Intelligence for Drug Toxicity and Safety. Vol. 40, Trends in Pharmacological Sciences. Elsevier Ltd; 2019. p. 624–635.

[16] Khan SR, Al Rijjal D, Piro A, Wheeler MB. Integration of AI and traditional medicine in drug discovery. Vol. 26, Drug Discovery Today. Elsevier Ltd; 2021. p. 982–992.

[17] Lavecchia A. Deep learning in drug discovery: opportunities, challenges and future prospects. Drug Discov Today [Internet]. 2019.

[18] Jiménez-Luna J, Grisoni F, Schneider G. Drug discovery with explainable artificial intelligence. Vol. 2, Nature Machine Intelligence. Nature Research; 2020. p. 573–584.

[19] Zhong F, Xing J, Li X, Liu X, Fu Z, Xiong Z, *et al*. Artificial intelligence in drug design. Vol. 61, Science China Life Sciences. Science in China Press; 2018. p. 1191–1204.

[20] Chen W, Liu X, Zhang S, Chen S. Artificial intelligence for drug discovery: Resources, methods, and applications. Mol Ther Nucleic Acids. 2023 Mar 14;31:691–702.

[21] Tripathi N, Goshisht MK, Sahu SK, Arora C. Applications of artificial intelligence to drug design and discovery in the big data era: a comprehensive review. Vol. 25, Molecular Diversity. Springer Science and Business Media Deutschland GmbH; 2021. p. 1643–1664.

[22] Panteleev J, Gao H, Jia L. Recent applications of machine learning in medicinal chemistry. Vol. 28, Bioorganic and Medicinal Chemistry Letters. Elsevier Ltd; 2018. p. 2807–2815.

[23] Gupta R, Srivastava D, Sahu M, Tiwari S, Ambasta RK, Kumar P. Artificial intelligence to deep learning: machine intelligence approach for drug discovery. Mol Divers. 2021 Aug 1;25(3):1315–1360.

[24] Wang L, Ding J, Pan L, Cao D, Jiang H, Ding X. Artificial intelligence facilitates drug design in the big data era. Vol. 194, Chemometrics and Intelligent Laboratory Systems. Elsevier B.V.; 2019.

[25] Mak KK, Pichika MR. Artificial intelligence in drug development: present status and future prospects. Vol. 24, Drug

Discovery Today. Elsevier Ltd; 2019. p. 773–80.

[26] Kulkov I. The role of artificial intelligence in business transformation: A case of pharmaceutical companies. Technol Soc. 2021 Aug 1;66:101629.

[27] Paul D, Sanap G, Shenoy S, Kalyane D, Kalia K, Tekade RK. Artificial intelligence in drug discovery and development. Vol. 26, Drug Discovery Today. Elsevier Ltd; 2021. p. 80–93.

[28] Klambauer G, Unterthiner T, Mayr A, Hochreiter S. Self-Normalizing Neural Networks. 2017.

[29] Kashyap K, Siddiqi MI. Recent advances in deep learning enabled approaches for identification of molecules of therapeutics relevance. Cheminformatics, QSAR and Machine Learning Applications for Novel Drug Development. 2023 Jan 1;503–518.

[30] Lahey SLJ, Rowley CN. Simulating protein-ligand binding with neural network potentials. Chem Sci. 2020 Mar 7;11(9):2362–2368.

[31] Lu J, Hou X, Wang C, Zhang Y. Incorporating Explicit Water Molecules and Ligand Conformation Stability in Machine-Learning Scoring Functions. J Chem Inf Model. 2019 Nov 25;59(11):4540–4549.

[32] Kolluri S, Lin J, Liu R, Zhang Y, Zhang W. Machine Learning and Artificial Intelligence in Pharmaceutical Research and Development: a Review. Vol. 24, AAPS Journal. Springer Science and Business Media Deutschland GmbH; 2022.

[33] Tripathi MK, Nath A, Singh TP, Ethayathulla AS, Kaur P. Evolving scenario of big data and Artificial Intelligence (AI) in drug discovery. Mol Divers. 2021 Aug 1;25(3):1439–1460.

[34] Chan HCS, Shan H, Dahoun T, Vogel H, Yuan S. Advancing Drug Discovery via Artificial Intelligence. Vol. 40, Trends in Pharmacological Sciences. Elsevier Ltd; 2019. p. 592–604.

[35] Green CP, Engkvist O, Pairaudeau G. The convergence of artificial intelligence and chemistry for improved drug discovery. Future Med Chem. 2018 Nov 1;10(22):2573–25766.

[36] Dara S, Dhamercherla S, Jadav SS, Babu CM, Ahsan MJ. Machine Learning in Drug Discovery: A Review. Artif Intell Rev. 2022 Mar 1;55(3): 1947–1999.

[37] Gallego V, Naveiro R, Roca C, Ríos Insua D, Campillo NE. AI in drug development: a multidisciplinary perspective. Mol Divers. 2021 Aug 1;25(3): 1461–1479.

[38] Patel V, Shah M. Artificial intelligence and machine learning in drug discovery and development. Vol. 2, Intelligent Medicine. Chinese Medical Association; 2022. p. 134–140.

[39] Bentwich I. Pharma's Bio-AI revolution. Drug Discov Today. 2023 May 1;28(5): 103515.

[40] Askin S, Burkhalter D, Calado G, El Dakrouni S. Artificial Intelligence Applied to clinical trials: opportunities and challenges. Health and Technology. Springer Science and Business Media Deutschland GmbH; 2023.

[41] Sandeep Ganesh G, Kolusu AS, Prasad K, Samudrala PK, Nemmani KVS. Advancing health care via artificial intelligence: From concept to clinic. Vol. 934, European Journal of Pharmacology. Elsevier B.V.; 2022.

[42] Lee CY, Chen YPP. Machine learning on adverse drug reactions for pharmacovigilance. Vol. 24, Drug Discovery Today. Elsevier Ltd; 2019. p. 1332–1343.

[43] Ding L, Liu C, Li Z, Wang Y. Incorporating Artificial Intelligence into Stroke Care and Research. Stroke. 2020 Dec 1;51(12):E351–354.

[44] Dankwa-Mullan I, Rivo M, Sepulveda M, Park Y, Snowdon J, Rhee K. Transforming Diabetes Care Through Artificial Intelligence: The Future Is Here. Popul Health Manag. 2019 Jun 1;22(3):229–242.

[45] Chaudhuri S, Long A, Zhang H, Monaghan C, Larkin JW, Kotanko P, *et al*. Artificial intelligence enabled applications in kidney disease. Vol. 34, Seminars in Dialysis. Blackwell Publishing Inc.; 2021. p. 5–16.

[46] Hassanzadeh P, Atyabi F, Dinarvand R. The significance of artificial intelligence in drug delivery system design. Vols 151–152, Advanced Drug Delivery Reviews. Elsevier B.V.; 2019. p. 169–190.

[47] Chumbita M, Cillóniz C, Puerta-Alcalde P, Moreno-García E, Sanjuan G, Garcia-Pouton N, *et al*. Can artificial intelligence improve the management of pneumonia. Vol. 9, Journal of Clinical Medicine. MDPI; 2020.

[48] Chakravarty K, Antontsev V, Bundey Y, Varshney J. Driving success in personalized medicine through AI-enabled computational modeling. Vol. 26, Drug Discovery Today. Elsevier Ltd; 2021. p. 1459–1465.

[49] Makne PD, Sontakke SS, Lakade RD, Tompe AS, Patil SS. ARTIFICIAL INTELLIGENCE: A REVIEW. World Journal of Pharmaceutical Research www.wjpr.net | [Internet]. 2015;12:739.

[50] Mehta CH, Narayan R, Nayak UY. Computational modeling for formulation design. Drug Discov Today. 2019 Mar 1;24(3):781–778.

[51] Khanna V, Ahuja R, Popli H. ROLE OF ARTIFICIAL INTELLIGENCE IN PHARMACEUTICAL MARKETING: A COMPREHENSIVE REVIEW [Internet]. Vol. 11, Journal of Advanced Scientific Research. 2020.

[52] Meshram DB. Artificial Intelligence in Pharma Industry-A Rising Concept Production and purification of Pharmaceutically Important Fibrinolytic Enzyme from Bacillus Species View project in silico characterization and Homology model building of Tau protein View project [Internet]. Available from: https://www.researchgate.net/publication/366371525

[53] Costa FF. Big data in biomedicine. Vol. 19, Drug Discovery Today. Elsevier Ltd; 2014. p. 433–40.

[54] Katsila T, Spyroulias GA, Patrinos GP, Matsoukas MT. Computational approaches in target identification and drug discovery.

Comput Struct Biotechnol J. 2016 Jan 1; 14:177–84.

[55] Aristodemou L, Tietze F. The state-of-the-art on Intellectual Property Analytics (IPA): A literature review on artificial intelligence, machine learning and deep learning methods for analysing intellectual property (IP) data. World Patent Information. 2018 Dec 1; 55: 37–51.

[56] Savoska S, Ristevski B. Towards implementation of big data concepts in a pharmaceutical company. Open Computer Science. 2020 Jan 1;10(1):343–356.

[57] Galetsi P, Katsaliaki K, Kumar S. Big data analytics in health sector: Theoretical framework, techniques and prospects. Vol. 50, International Journal of Information Management. Elsevier Ltd; 2020. p. 206–216.

[58] Mehta N, Pandit A, Shukla S. Transforming healthcare with big data analytics and artificial intelligence: A systematic mapping study. Vol. 100, Journal of Biomedical Informatics. Academic Press Inc.; 2019.

[59] Brown N, Cambruzzi J, Cox PJ, Davies M, Dunbar J, Plumbley D, *et al*. Big Data in Drug Discovery. In: Progress in Medicinal Chemistry. Elsevier B.V.; 2018. p. 277–356.

[60] Zhao L, Ciallella HL, Aleksunes LM, Zhu H. Advancing computer-aided drug discovery (CADD) by big data and data-driven machine learning modeling. Vol. 25, Drug Discovery Today. Elsevier Ltd; 2020. p. 1624–1638.

[61] Zhang L, Tan J, Han D, Zhu H. From machine learning to deep learning: progress in machine intelligence for rational drug discovery. Vol. 22, Drug Discovery Today. Elsevier Ltd; 2017. p. 1680–1685.

[62] Dossetter AG, Ecker G, Laverty H, Overington J. 'Big data' in pharmaceutical science: Challenges and opportunities. Vol. 6, Future Medicinal Chemistry. Future Science; 2014. p. 857–864.

[63] Furstenau LB, Leivas P, Sott MK, Dohan MS, López-Robles JR, Cobo MJ, *et al*. Big data in healthcare: Conceptual network structure, key challenges and opportunities. Digital Communications and Networks. 2023.

[64] Jiao Z, Ji H, Yan J, Qi X. Application of big data and artificial intelligence in epidemic surveillance and containment. Intelligent Medicine. 2023 Feb 1;3(1):36–43.

[65] Schuhmacher A, Gatto A, Kuss M, Gassmann O, Hinder M. Big Techs and startups in pharmaceutical R&D – A 2020 perspective on artificial intelligence. Vol. 26, Drug Discovery Today. Elsevier Ltd; 2021. p. 2226–2231.

[66] Vijayan RSK, Kihlberg J, Cross JB, Poongavanam V. Enhancing preclinical drug discovery with artificial intelligence. Drug Discov Today. 2022 Apr, 27(4):967–984.

[67] Muehlematter UJ, Daniore P, Vokinger KN. Approval of artificial intelligence and machine learning-based medical devices in the USA and Europe (2015–20): a comparative analysis. Lancet Digit Health. 2021 Mar, 3(3): 195–203.

[68] Capecchi A, Awale M, Probst D, Reymond JL. PubChem and ChEMBL beyond Lipinski [Internet]. Available from: https://chemaxon.com

[69] Tormay P. Big Data in Pharmaceutical R&D: Creating a Sustainable R&D Engine. Pharmaceut Med. 2015, 29(2):87–92.

[70] Hird N, Ghosh S, Kitano H. Digital health revolution: Perfect storm or perfect opportunity for pharmaceutical R&D? Vol. 21, Drug Discovery Today. Elsevier Ltd; 2016. p. 900–911.