



Espailat. A International Journal of Computer Science and Mobile Applications, Vol. 12 Issue 1, January-2024, pg. 01-16.

ISSN: 2321-8363

Impact Factor: 6.308

(An Open Accessible, Fully Refereed and Peer Reviewed Journal)

Enhancing Drug Discovery and Development through Artificial Intelligence

Alejandro Espailat

M.D. South Florida Eye Institute, Fort Lauderdale, Florida, US

E-mail: drespailat@evnmiami.com

Received date: 23 May 2024, Manuscript No. IJCSMA-24-136856, **Editor assigned:** 26 May 2024, Pre QC No. IJCSMA-24-136856 (PQ); **Reviewed:** 29 May 2024, QC No. IJCSMA-24-136856 (Q); **Revised:** 1 June 2024, Manuscript No. IJCSMA-24-136856 (R); **Published date:** 14 June 2024; DOI. 10.5281/zenodo.8116501

Abstract

Artificial Intelligence (AI) has emerged as a powerful tool in drug discovery and development, revolutionizing how new medicines are identified, optimized, and brought to market. Several drug discovery processes, including 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, have leveraged AI to achieve their goals. This review provides an in-depth analysis of AI's various applications in the pharmaceutical industry, including virtual screening, molecular modeling, target identification, and drug repurposing. We discuss the challenges and opportunities associated with AI in drug discovery and its impact on the future of medicine.

Keywords: Drug discovery; Artificial intelligence; Machine learning; Deep learning; Virtual screening; Molecular modelling; Target identification; Drug repurposing

©2024, IJCSMA All Rights Reserved, www.ijcsma.com



This work is licensed under a [Creative Commons Attribution 4.0 International License](https://creativecommons.org/licenses/by/4.0/).



Espaillet. A International Journal of Computer Science and Mobile Applications, Vol. 12 Issue 1, January-2024, pg. 01-16.

ISSN: 2321-8363

Impact Factor: 6.308

(An Open Accessible, Fully Refereed and Peer Reviewed Journal)

1. Introduction

Drug development consists of four key phases: drug discovery, pre-clinical research, clinical development and market approval (Phase I focuses on evaluating pharmacokinetics, safety, and tolerability in healthy volunteers. Phase II involves testing efficacy and dose response in a small cohort group of patients with the target disease. Phase III comprises large-scale studies to verify safety and effectiveness.), and post-approval surveillance (Phase IV). The initial drug discovery stage involves identifying and creating new chemical compounds that target specific protein structures related to diseases and medical conditions. Drug design is vital in drug discovery, refining potential drug compounds through lead optimization [1].

Introducing AI techniques in drug discovery and development processes shows potential for accelerating timelines and reducing costs [2]. The substantial financial and time investments required in bringing a new drug to market, with an estimated average price of 2.6 billion USD and a timeline exceeding ten years, illustrate the significance of streamlining drug development efforts. Additionally, the low success rate of new therapeutic agents reaching the market from Phase I clinical trials, estimated at less than 10%, highlights the challenges faced in this field [3-7]. The initial phase of the drug discovery process centers on identifying pertinent targets, such as specific genes and proteins linked to disease pathways, followed by the quest for suitable pharmaceutical compounds or drug analogs capable of interacting with these targets [8, 9]. The concept of big data pertains to vast datasets beyond the scope of conventional data analysis tools and techniques, owing to their substantial size, rapid data generation rates, and diverse data modalities. The abundance of large-scale biomedical data repositories is a valuable resource in facilitating these endeavors. Concurrently, the evolution of AI technologies has streamlined data analysis practices, enabling the application of various Machine Learning (ML) methodologies to explore, interpret, and discern essential insights, patterns, and relationships within extensive biomedical datasets [10]. Researchers can use ML, Deep Learning (DL), and other AI tools to pinpoint promising drug candidates, forecast their potential efficacy and safety profiles, and refine their molecular structures to amplify therapeutic potency [11].

ML algorithms push forward drug discovery, providing significant benefits to pharmaceutical companies. These algorithms have been instrumental in creating predictive models for evaluating chemical, biological, and physical properties of compounds in drug development [12-14]. ML algorithms are versatile tools that can be integrated into all stages of the drug discovery process. They have been utilized to uncover new applications for existing drugs, forecast interactions between drugs and proteins, assess drug efficacy, determine safety markers, and enhance the bioactivity of molecules [15-18]. Commonly used ML algorithms in drug discovery include Random Forest (RF), Naive Bayesian (NB), and Support Vector Machine (SVM) [19-21].

This review article provides a comprehensive overview of the critical applications of AI in drug discovery and development. We discuss the use of AI in virtual screening, which involves rapidly screening large compound libraries to identify potential drug candidates [5, 22-24]. We also explore how AI is being utilized in molecular

©2024, IJCSMA All Rights Reserved, www.ijcsma.com



This work is licensed under a [Creative Commons Attribution 4.0 International License](https://creativecommons.org/licenses/by/4.0/).



Espaillet. A International Journal of Computer Science and Mobile Applications, Vol. 12 Issue 1, January-2024, pg. 01-16.

ISSN: 2321-8363

Impact Factor: 6.308

(An Open Accessible, Fully Refereed and Peer Reviewed Journal)

modeling to predict the binding affinity of drug candidates to their target proteins, as well as in target identification to identify novel therapeutic targets for specific diseases. Additionally, we examine the growing trend of drug repurposing, where AI is used to discover new indications for existing drugs.

2. Artificial Intelligence Fundamentals

AI is a branch of computer science that focuses on developing machine systems capable of performing tasks that typically require human intelligence. AI systems are designed to learn from data, identify patterns, make decisions, and solve complex problems [25]. Various techniques and approaches are used in AI, such as ML, DL, Natural Language Processing (NLP), and computer vision [26, 27]. ML algorithms enable AI systems to improve their performance over time by learning from data without being explicitly programmed. DL, a subset of ML, uses neural networks to process large amounts of data and extract meaningful patterns [28]. It is important to note that ML algorithms are not uniform within AI. There are two primary categories of ML algorithms: supervised and unsupervised learning. Supervised learning involves training with labeled data to predict labels for new samples, while unsupervised learning identifies patterns in unlabeled data. Unsupervised learning often transforms the data into a lower-dimensional space to facilitate pattern recognition when working with high-dimensional data. This dimension reduction enhances efficiency and aids in the interpretation of patterns. Additionally, the fusion of supervised and unsupervised learning occurs in semi-supervised and reinforcement learning approaches, offering flexibility for diverse datasets.

The success of ML algorithms in drug discovery heavily relies on the availability of vast quantities of high-quality data and well-defined training sets. This requirement is particularly crucial in precision medicine, where a meticulous understanding of various pan-omic data types (such as genomic, transcriptomic, and proteomic) is essential for tailoring effective personalized therapies. Characteristics such as known data and thorough training sets are vital for the development, refinement, and efficacy of ML algorithms supporting drug discovery in the modern era [28, 29]. The principles related to drug discovery and computer-assisted drug design methods can be found in the Computer-Assisted Drug Design [29, 30].

3. Virtual Screening for Drug Discovery

Drug discovery projects typically commence when there is a lack of effective drugs for a specific disease or when existing treatments exhibit limited efficacy or substantial side effects [31]. The initial phase involves formulating a hypothesis that manipulating a particular target, such as an enzyme or receptor, will result in therapeutic benefits for the disease. This process includes identifying and validating the target. Subsequently, rigorous assays are conducted to identify potential compounds (hits) and develop them into potential drug candidates (leads) through hit discovery, hit-to-lead transformation, and lead optimization stages. These candidates then undergo preclinical testing and clinical trials. Successful candidates can eventually be approved and marketed as medical treatments for the targeted disease in **Figure 1**.

©2024, IJCSMA All Rights Reserved, www.ijcsma.com



This work is licensed under a [Creative Commons Attribution 4.0 International License](https://creativecommons.org/licenses/by/4.0/).



(An Open Accessible, Fully Refereed and Peer Reviewed Journal)

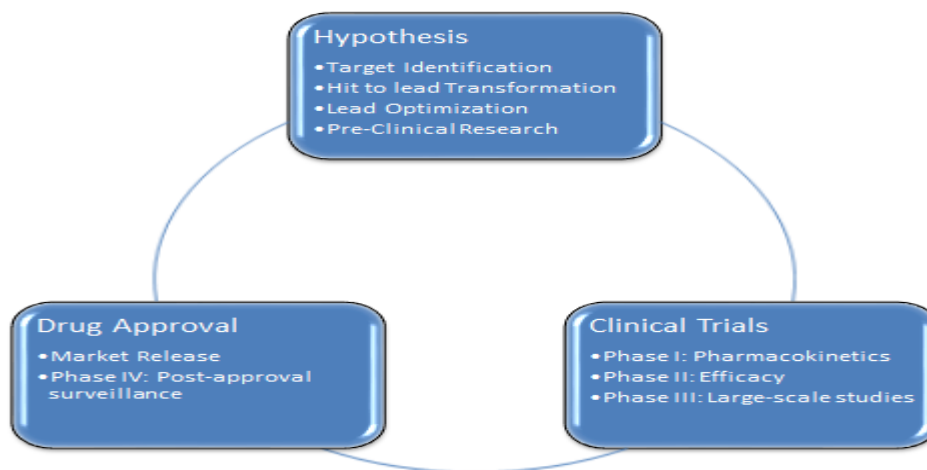


Figure 1. Drug Discovery Process Pipeline.

Since the 1980s, High-Throughput Screening (HTS) has been utilized to expedite the discovery of small-molecule drugs, enhancing efficiency by leveraging automation and large chemical libraries [32, 33]. HTS generates significant Structure-Activity Relationship (SAR) datasets, which enrich chemical databases like PubChem and ZINC [34, 35]. AI Virtual Screening (VS) involves computational methods to sift through these vast chemical libraries for potentially active compounds for further in vitro and in vivo testing, relying on knowledge of the target (structure-based VS) or known active molecules (ligand-based VS) [36]. This method aims to improve the speed of identification of active molecules based on the hypothesis that targeting a specific molecule can treat a disease, utilizing agonists and antagonists as major classes of drugs with different mechanisms of action. Agonists activate the target to evoke a biological response, while antagonists bind to the target to block this response [37]. Quantifying activity involves measuring affinity (or potency) and efficacy. Affinity reflects how strongly a molecule binds to a target, potency indicates the amount needed to elicit a specific effect, and efficacy describes the magnitude of the impact, such as inhibiting an enzyme by a particular percentage amount.

An ideal drug candidate needs to have sufficient activity as a ligand and exhibit binding specificity to specific targets to avoid unexpected side effects [38]. High selectivity is desired to prevent binding to multiple targets [39]. A combination of criteria, including physicochemical, pharmacokinetic, and pharmacodynamic properties, must be considered for drug candidates. Other properties like the Synthetic Accessibility Score (SAS) and the Quantitative Estimation of Drug-likeness (QED) are also crucial during compound synthesis. Based on the molecular fragments involved, the SAS rating system indicates the complexity of synthesizing a molecule, ranging from 1 (easy) to 10 (difficult). The QED is a measure, on a scale of 0 to 1, that predicts the likelihood that a molecule is a suitable drug





Espailat. A International Journal of Computer Science and Mobile Applications, Vol. 12 Issue 1, January-2024, pg. 01-16.

ISSN: 2321-8363

Impact Factor: 6.308

(An Open Accessible, Fully Refereed and Peer Reviewed Journal)

candidate.

Drug discovery involves a multi-objective optimization with predictive models like Quantitative Structure-Activity Relationship (QSAR) modeling used to map molecular structure to property values. QSAR models can also be used in reverse to identify structural features for optimal properties and guide drug design from scratch, known as de novo drug design [40].

Drug design goes beyond screening existing chemical libraries to explore the vast chemical space, comprising all potential small molecules [41, 42]. This space is estimated to be extensive and involves a continuous cycle of Design-Making-Testing-Analysis (DMTA) which includes iterative organic synthesis and property testing [43]. To effectively navigate this chemical space, quantitative drug design has been proposed since the late 1970s [44]. Drug design is centered on two main questions: whether molecular properties can be inferred from molecular structures and which structural characteristics are relevant for specific molecular properties. The former question forms the basis of VS, while QSAR models address the latter. Drug design can be seen as an extension of virtual screening, encompassing tasks such as predicting molecular properties and generating molecules, which are essential components of current AI-driven drug discovery processes [45].

4. Molecular Modeling for De Novo Drug Design

Calculating the affinity between a ligand and a biological target is crucial in drug discovery. Various virtual screening techniques can quickly classify millions of compounds as 'active' or 'inactive' for a specific target. Understanding the binding affinity and interaction strength between a ligand and target protein is a vital aspect of drug discovery, as it aids in identifying potential drug candidates. Computational methods for predicting binding affinity offer significant time and cost savings compared to traditional laboratory experiments [46]. In the initial stages of small molecule drug discovery, rational drug design targets specific proteins, with potential molecules chosen as hits based on their binding affinity. While numerous computational methods have been developed for predicting protein-ligand binding affinity, they often rely on simplifying assumptions that lead to inaccuracies and high false-positive rates in hit identification [47-52]. Additionally, docking scores for hit prioritization during virtual screening may provide unreliable information. Therefore, accurately predicting binding affinity and identifying target hits through structure-based methods remain challenging. Using ML approaches involves training models with experimental protein-ligand data to predict the binding affinity of new protein-ligand complexes. However, challenges persist in accurately representing protein-ligand interactions, accounting for protein flexibility, selecting appropriate descriptors, and dealing with the wide range of ligand affinity values [53]. Among ML models, the RF method has shown improved affinity prediction compared to other techniques, with models like RF-Score and SFCscoreRF being developed.

DL models are gaining popularity for their high performance in various fields, including drug discovery and visual and speech recognition [54, 55]. Several DL methods have been proposed for predicting ligand affinity and designing new drugs. For example, DEELIG, a DL model by Ahmed *et al.* employs Convolutional Neural Networks

©2024, IJCSMA All Rights Reserved, www.ijcsma.com



This work is licensed under a [Creative Commons Attribution 4.0 International License](https://creativecommons.org/licenses/by/4.0/).



Espailat. A International Journal of Computer Science and Mobile Applications, Vol. 12 Issue 1, January-2024, pg. 01-16.

ISSN: 2321-8363

Impact Factor: 6.308

(An Open Accessible, Fully Refereed and Peer Reviewed Journal)

(CNN) to identify spatial relationships within data, using a 3D grid of atoms to represent protein-ligand complexes. Li *et al.* developed Deep Atom based on a CNN that extracts atom interaction features from the voxelized structure of protein-ligand complexes [56, 57].

Additionally, Jiménez *et al.* introduced KDEEP, which uses a CNN-based model utilizing 3D voxel representations of proteins and ligands [58]. Limbu *et al.* introduced a novel Hybrid Neural Network (HNN) DL model that includes the 'HNN-Lenovo' and 'HNN-affinity' methods [46]. These methods leverage distinct learning frameworks to enhance de novo drug design prediction accuracy.

5. Target Identification for Drug Discovery

Modeling diseases and identifying targets are vital early stages in drug discovery, and they play a crucial role in determining the success of drug development. Target identification can be categorized into experimental, multi-omic, and computational strategies. Combined, these methods can create innovative therapeutic ideas in preliminary target identification, leading to a deeper comprehension of intricate diseases [59].

5.1. Experimental Target Identification:

- Affinity-based biochemical methods: Small-molecule affinity probes stand out for their ability to label proteins without leaving a trace when interacting with their ligands [60].
- Comparative profiling: A widely used quantitative proteomics technique called Stable Isotope Labeling by Amino Acids in Cell Culture (SILAC) utilizes stable isotope-labeled amino acids to distinguish between cellular proteomes accurately [61]. Various studies across different cancer types, such as hepatocellular carcinoma, multiple myeloma, endometrial cancer, and colorectal cancer, have shown SILAC's effectiveness in identifying critical factors in disease development [62-65].
- Chemical/genetic screening: Chemical/genetic screening methods, like RNA interference (RNAi) and CRISPR-Cas9 gene editing, have long fascinated biologists due to their specificity and efficiency. CRISPR technology has significantly advanced our understanding of molecular and pharmacological aspects of human diseases [66].

5.2. Multiomic Target Identification:

Multiomic data offers researchers a comprehensive view of molecular information from various sources, encompassing both stable genomic data and dynamic expression and metabolic profiles across space and

©2024, IJCSMA All Rights Reserved, www.ijcsma.com



This work is licensed under a [Creative Commons Attribution 4.0 International License](https://creativecommons.org/licenses/by/4.0/).



Espailat. A International Journal of Computer Science and Mobile Applications, Vol. 12 Issue 1, January-2024, pg. 01-16.

ISSN: 2321-8363

Impact Factor: 6.308

(An Open Accessible, Fully Refereed and Peer Reviewed Journal)

time [67]. Genomics, the oldest and most established omics field, primarily focuses on genetic variations within the DNA sequence [68]. Through large-scale Genome-Wide Association Studies (GWAS) driven by advanced sequencing technologies, numerous links have been discovered between genetic variations and complex diseases or traits leading to groundbreaking therapeutic advancements like the development of cystic fibrosis modulator drugs targeting CFTR mutations and new treatments for inflammatory bowel disease that target the disease-related gene IL23A [69, 70]. Furthermore, recent meta-analyses of extensive GWAS data have unveiled previously unknown genetic loci associated with various diseases, creating opportunities for drug repurposing [71].

5.3. Computational Target Identification:

- Traditional experimental methods for identifying targets are time-consuming and resource-intensive, prompting the exploration of computational techniques as efficient alternatives. Various approaches, such as pharmacophore screening, reverse docking and structure similarity analysis predict potential biological targets for small molecules based on factors like protein structure and compound chemical structure [72-74]. In pharmaceutical research, DL techniques, including generative adversarial networks and recurrent neural networks, have gained substantial traction due to their ability to process data and extract features through multiple layers of nodes [75]. These advanced algorithms have found applications in small-molecule design aging studies and predicting drug responses from gene expression data [76-78]. By leveraging diverse data sources and text analysis, DL methods address pressing medical challenges, including severe and unmet medical conditions.

Moreover, advanced language models are crucial in expedited biomedical text mining for discovering therapeutic targets. These models, like BioGPT by Microsoft and ChatPandaGPT by Insilico Medicine, use extensive training on vast text data to connect diseases, genes, and biological processes [79]. They facilitate the quick identification of disease mechanisms, potential drug targets, and biomarkers. While these language models excel at understanding complex scientific concepts and accelerating disease hypothesis generation, they may unknowingly perpetuate human biases and need more discernment to validate input data accuracy. Additionally, their reliance on published information might restrict their ability to uncover innovative targets. Therefore, it is recommended to acknowledge these model limitations and complement their use with other approaches to ensure the discovery of genuinely novel therapeutic targets.

Using AI-generated synthetic data is beneficial for target identification in medical research. Synthetic data, created by AI algorithms to mimic real-world patterns, can help researchers explore a broader range of scenarios, especially in areas with limited experimental data like rare diseases. By generating synthetic data based on existing knowledge, AI can reveal potential therapeutic targets that may have been overlooked, aiding in the validation of predictions and addressing data imbalance issues [80-82]. However, it's essential to acknowledge the limitations of synthetic data, as models may not capture unknown complexities, and ethical concerns could arise from simulating under-represented

©2024, IJCSMA All Rights Reserved, www.ijcsma.com



This work is licensed under a [Creative Commons Attribution 4.0 International License](https://creativecommons.org/licenses/by/4.0/).



Espailat. A International Journal of Computer Science and Mobile Applications, Vol. 12 Issue 1, January-2024, pg. 01-16.

ISSN: 2321-8363

Impact Factor: 6.308

(An Open Accessible, Fully Refereed and Peer Reviewed Journal)

populations. Robust validation and quality control measures are crucial to ensuring the reliability and relevance of AI-generated synthetic data in biomedical research [83-85].

6. Drug Repurposing for Drug Discovery

Drug repurposing, a feasible and promising approach, has garnered increasing interest from governments and pharmaceutical companies for its excellent track record in saving time and money. It involves identifying new medical uses for existing drugs initially developed for different purposes. This approach presents a quicker and more cost-efficient way to create new treatments [86, 87].

Drug repurposing strategies can be categorized into drug-based and disease-based approaches. The familiar premise is that a drug may effectively treat multiple diseases with similarities or interconnections [88, 89]. Target associations can be complex due to the multifunctional nature of drugs, and computational drug repurposing faces challenges in distinguishing drug targets from other gene products indirectly involved in target activity [88]. Traditional methods may need more datasets and environmental variations, potentially leading to inaccuracies. Still, the growing biomedical and pharmaceutical data has improved computational approaches, such as data mining and ML, to repurpose drugs better [90, 91]. These advanced methods help uncover therapeutic opportunities by analyzing interactions among biological entities like genes, proteins, drugs, and diseases within complex networks. Some ML techniques for repurposing drugs include k-nearest neighbors, RF and SVM [92-94].

Various studies have used ML techniques like collaborative filtering to predict new drug-disease associations based on gene expression patterns [95, 96]. One study created drug similarity datasets to identify potential repurposed uses for FDA-approved drugs by analyzing known and unknown associations [97]. They used SVM classification and collaborative filtering to predict novel drug-disease links. Another study proposed a computational framework integrating different data sources to predict similarities between drugs, diseases, and drug-disease pairs [98]. They used strategies like block coordinate descent and causal inference-probabilistic matrix factorization to classify new drug-disease associations. These approaches help identify potential off-target drug interactions and discover new connections within the drug-disease network.

Due to their ability to uncover intricate patterns, DL models are well-suited for analyzing complex data such as electronic health records, the entire proteome, and the human genome [99]. This makes them particularly applicable in life sciences fields like drug repurposing. DL stands out from traditional ML techniques for its neural network flexibility [100]. DL methods offer various advantages over conventional approaches, particularly their ability to automatically model and learn complex features, and have proven beneficial in identifying repurposed drugs. Nonetheless, several challenges persist, including the “black box phenomena” and opaque nature of DL models, a lack of interpretability, output reliance on input data, and the necessity of extensive standardized biochemical datasets to achieve optimal learning and performance outcomes. Despite the availability of vast amounts of data, the process of extracting and preparing standardized data for ML and DL applications remains a significant challenge. It

©2024, IJCSMA All Rights Reserved, www.ijcsma.com



This work is licensed under a [Creative Commons Attribution 4.0 International License](https://creativecommons.org/licenses/by/4.0/).



Espailat. A International Journal of Computer Science and Mobile Applications, Vol. 12 Issue 1, January-2024, pg. 01-16.

ISSN: 2321-8363

Impact Factor: 6.308

(An Open Accessible, Fully Refereed and Peer Reviewed Journal)

is hoped that the development of extensive datasets in the future will facilitate the establishment of standardized DL models for drug repurposing [101].

7. Conclusion

In conclusion, AI has the potential to revolutionize the field of drug discovery and development, transforming the way medicines are discovered, optimized, and brought to market. By harnessing the power of ML and other AI technologies, researchers can accelerate the drug development process, reduce costs, and improve the success rates of new therapeutic interventions. As AI continues to evolve and mature, its integration into the pharmaceutical industry will be crucial in advancing personalized medicine and improving healthcare outcomes for patients worldwide.

References

1. Tamimi, Nihad AM, & Peter Ellis. "Drug development: from concept to marketing!." *Nephron Clin Pract.* 113.3 (2009): c125-c131.
2. Sarkar, Chayna. et al. "Artificial intelligence and machine learning technology driven modern drug discovery and development." *Int J Mol Sci.* 24.3 (2023): 2026.
3. Paul, D. et al. "Artificial Intelligence in Drug Discovery and Development." *Drug Discov Today.* 2021;26(1):80-93.
4. Gupta, Rohan., et al. "Artificial intelligence to deep learning: machine intelligence approach for drug discovery." *Mol Divers.* 25 (2021): 1315-1360.
5. Deng, Jianyuan., et al. "Artificial intelligence in drug discovery: applications and techniques." *Brief Bioinform.* 23.1 (2022): bbab430.
6. Mullard, Asher. "New drugs cost US \$2.6 billion to develop." *Nat Rev Drug Discov.* 13.12 (2014): 877-877.
7. Dowden, Helen, & Jamie Munro. "Trends in clinical success rates and therapeutic focus." *Nat Rev Drug Discov* 18.7 (2019): 495-496.
8. Gandomi, Amir, & Murtaza Haider. "Beyond the hype: Big data concepts, methods, and analytics." *Int J Inf Manag* 35.2 (2015): 137-144.

©2024, IJCSMA All Rights Reserved, www.ijcsma.com



This work is licensed under a [Creative Commons Attribution 4.0 International License](https://creativecommons.org/licenses/by/4.0/).



Espailat. A International Journal of Computer Science and Mobile Applications, Vol. 12 Issue 1, January-2024, pg. 01-16.

ISSN: 2321-8363

Impact Factor: 6.308

(An Open Accessible, Fully Refereed and Peer Reviewed Journal)

9. Brazma, Alvis., et al. "[20] "Data Storage and Analysis in ArrayExpress" *Methods Enzymol.* 411 (2006): 370-386.
10. van IJzendoorn, David GP, et al. "Machine learning analysis of gene expression data reveals novel diagnostic and prognostic biomarkers and identifies therapeutic targets for soft tissue sarcomas." *PLoS Comput Biol.* 15.2 (2019): e1006826.
11. Parvatikar, Prachi P., et al. "Artificial intelligence: Machine learning approach for screening large database and drug discovery." *Antivir Res.* (2023): 105740.
12. Lo, Yu-Chen., et al. "Machine learning in chemoinformatics and drug discovery." *Drug Discov Today* 23.8 (2018): 1538-1546.
13. Talevi, Alan., et al. "Machine learning in drug discovery and development part 1: a primer." *CPT: Pharmacomet Syst Pharmacol.* 9.3 (2020): 129-142.
14. Gertrudes, J. Castro., et al. "Machine learning techniques and drug design." *Curr Med Chem.* 19.25 (2012): 4289-4297.
15. Khamis, Mohamed A., et al. "Machine learning in computational docking." *Artif Intell Med.* 63.3 (2015): 135-152.
16. Leelananda, Sumudu P., & Steffen Lindert. "Computational methods in drug discovery." *Beilstein J Org Chem.* 12.1 (2016): 2694-2718.
17. Maia, Eduardo Habib Bechelane, et al. "Structure-based virtual screening: from classical to artificial intelligence." *Front Chem.* 8 (2020): 343.
18. Usha, Talambedu, et al. "Recent updates on computer-aided drug discovery: time for a paradigm shift." *Curr Top Med Chem.* 17.30 (2017): 3296-3307.
19. Breiman, Leo. "Random forests." *Mach Learn.* 45 (2001): 5-32.
20. Webb, Geoffrey I., & Claude Sammut, eds. *Encyclopedia of machine learning.* Springer, 2010.
21. Cortes, Corinna & Vladimir Vapnik. "Support-vector networks." *Mach Learn* 20 (1995): 273-297.
22. Murugan, N. Arul, et al. "Artificial intelligence in virtual screening: Models versus experiments." *Drug Discov. Today* 27.7 (2022): 1913-1923.
23. Gentile, Francesco, et al. "Artificial intelligence-enabled virtual screening of ultra-large chemical libraries with deep docking." *Nat Protoc.* 17.3 (2022): 672-697.

©2024, IJCSMA All Rights Reserved, www.ijcsma.com



This work is licensed under a [Creative Commons Attribution 4.0 International License](https://creativecommons.org/licenses/by/4.0/).



Espailat. A International Journal of Computer Science and Mobile Applications, Vol. 12 Issue 1, January-2024, pg. 01-16.

ISSN: 2321-8363

Impact Factor: 6.308

(An Open Accessible, Fully Refereed and Peer Reviewed Journal)

24. Stumpfe, Dagmar, & Jürgen Bajorath. "Current trends, overlooked issues, and unmet challenges in virtual screening." *J Chem Inf Model.* 60.9 (2020): 4112-4115.
25. Bhinder, Bhavneet, et al. "Artificial intelligence in cancer research and precision medicine." *Cancer Discov.* 11.4 (2021): 900-915.
26. Choi, Rene Y., et al. "Introduction to machine learning, neural networks, and deep learning." *Transl Vis Sci Technol.* 9.2 (2020): 14-14.
27. Patel, Lauv, et al. "Machine learning methods in drug discovery." *Molecules* 25.22 (2020): 5277.
28. Haug, Charlotte J., & Jeffrey M. Drazen. "Artificial intelligence and machine learning in clinical medicine, 2023." *N Engl J Med.* 388.13 (2023): 1201-1208.
29. Trigg, David J., & John B. Taylor, eds. *Comprehensive medicinal chemistry II.* Vol. 8. Elsevier, 2006.
30. Duch, Wlodzislaw, et al. "Artificial intelligence approaches for rational drug design and discovery." *Curr Pharm Des.* 13.14 (2007): 1497-1508.
31. Hughes, James P., et al. "Principles of early drug discovery." *Br J Pharmacol.* 162.6 (2011): 1239-1249.
32. Pereira, D. A., & J. A. Williams. "Origin and evolution of high throughput screening." *Br J Pharmacol.* 152.1 (2007): 53-61.
33. Bender, Andreas, et al. "Which aspects of HTS are empirically correlated with downstream success?." *Curr Opin Drug Discov Dev.* 11.3 (2008): 327.
34. Wang, Yanli, et al. "Pubchem bioassay: 2017 update." *Nucleic Acids Res.* 45.D1 (2017): D955-D963.
35. Sterling, Teague, & John J. Irwin. "ZINC 15—ligand discovery for everyone." *J Chem Inf Model.* 55.11 (2015): 2324-2337.
36. Scior, Thomas, et al. "Recognizing pitfalls in virtual screening: a critical review." *J Chem Inf Model.* 52.4 (2012): 867-881.
37. Salahudeen, Mohammed Saji, & Prasad S. Nishtala. "An overview of pharmacodynamic modelling, ligand-binding approach and its application in clinical practice." *Saudi Pharm J.* 25.2 (2017): 165-175.
38. Hu, Ye, & Jürgen Bajorath. "Compound promiscuity: what can we learn from current data?" *Drug Discov Today* 18.13-14 (2013): 644-650.

©2024, IJCSMA All Rights Reserved, www.ijcsma.com



This work is licensed under a [Creative Commons Attribution 4.0 International License](https://creativecommons.org/licenses/by/4.0/).



Espailat. A International Journal of Computer Science and Mobile Applications, Vol. 12 Issue 1, January-2024, pg. 01-16.

ISSN: 2321-8363

Impact Factor: 6.308

(An Open Accessible, Fully Refereed and Peer Reviewed Journal)

39. Yusof, Iskander, et al. "Finding the rules for successful drug optimisation." *Drug Discov Today* 19.5 (2014): 680-687.
40. Nicolaou, Christos A., & Nathan Brown. "Multi-objective optimization methods in drug design." *Drug Discov Today: Technologies* 10.3 (2013): e427-e435.
41. Schneider, Gisbert, and Uli Fechner. "Computer-based de novo design of drug-like molecules." *Nat Rev Drug Discov* 4.8 (2005): 649-663.
42. Dobson, Christopher M. "Chemical space and biology." *Nat* 432.7019 (2004).
43. Schneider, Gisbert. "Automating drug discovery." *Nat Rev Drug Discov* 17.2 (2018): 97-113.
44. Van Drie, John H. "Computer-aided drug design: the next 20 years." *J Comput-Aided Mol Des.* 21.10 (2007): 591-601.
45. Jiménez-Luna, José, et al. "Drug discovery with explainable artificial intelligence." *Nat Mach Intell.* 2.10 (2020): 573-584.
46. Limbu, Sarita, & Sivanesan Dakshanamurthy. "A New Hybrid Neural Network Deep Learning Method for Protein–Ligand Binding Affinity Prediction and De Novo Drug Design." *Int J Mol Sci.* 23.22 (2022): 13912.
47. Ballester, Pedro J., & John BO Mitchell. "A machine learning approach to predicting protein–ligand binding affinity with applications to molecular docking." *Bioinformatics* 26.9 (2010): 1169-1175.
48. Kitchen, Douglas B., et al. "Docking and scoring in virtual screening for drug discovery: methods and applications." *Nat Rev Drug discov* 3.11 (2004): 935-949.
49. Leach, Andrew R., et al. "Prediction of protein– ligand interactions. Docking and scoring: successes and gaps." *J Med Chem.* 49.20 (2006): 5851-5855.
50. Moitessier, N., et al. "Towards the development of universal, fast and highly accurate docking/scoring methods: a long way to go." *Br J Pharmacol* 153.S1 (2008): S7-S26.
51. Huang, Niu, et al. "Molecular mechanics methods for predicting protein–ligand binding." *Phys Chem Chem Phys* 8.44 (2006): 5166-5177.
52. Irwin, John J. "Community benchmarks for virtual screening." *J Comput-Aided Mol Des.* 22 (2008): 193-199.
53. Li, Hongjian, et al. "Improving AutoDock Vina using random forest: the growing accuracy of binding

©2024, IJCSMA All Rights Reserved, www.ijcsma.com



This work is licensed under a [Creative Commons Attribution 4.0 International License](https://creativecommons.org/licenses/by/4.0/).



Espailat. A International Journal of Computer Science and Mobile Applications, Vol. 12 Issue 1, January-2024, pg. 01-16.

ISSN: 2321-8363

Impact Factor: 6.308

(An Open Accessible, Fully Refereed and Peer Reviewed Journal)

affinity prediction by the effective exploitation of larger data sets." *Mol Inform* 34.2-3 (2015): 115-126.

54. Lin, Tsung-Yu, et al. "Bilinear CNN models for fine-grained visual recognition." *Proc IEEE Int Conf Comput Vis*. 2015.
55. Pereira, Janaina Cruz, et al. "Boosting docking-based virtual screening with deep learning." *J Chem Inf Model*. 56.12 (2016): 2495-2506.
56. Ahmed, Asad, et al. "DEELIG: A deep learning approach to predict protein-ligand binding affinity." *Bioinform Biol Insights* 15 (2021): 11779322211030364.
57. Li, Yanjun, et al. "DeepAtom: A framework for protein-ligand binding affinity prediction." *2019 IEEE Int Conf Bioinform Biomed. (BIBM)*. IEEE, 2019.
58. Jiménez, José, et al. "K deep: protein–ligand absolute binding affinity prediction via 3d-convolutional neural networks." *J Chem Inf Model*. 58.2 (2018): 287-296.
59. Pun, Frank W., et al. "AI-powered therapeutic target discovery." *Trends Pharmacol Sci*. (2023).
60. Shiraiwa, Kazuki, et al. "Chemical tools for endogenous protein labeling and profiling." *Cell Chem Biol*. 27.8 (2020): 970-985.
61. Ong, Shao-En, & Matthias Mann. "A practical recipe for stable isotope labeling by amino acids in cell culture (SILAC)." *Nat Protoc*. 1.6 (2006): 2650-2660.
62. Jin, Jing, et al. "SIRT3-dependent deacetylation of cyclin E2 prevents hepatocellular carcinoma growth." *EMBO Rep* 24.5 (2023): e56052.
63. Li, Xin, et al. "Deacetylation induced nuclear condensation of HP1 γ promotes multiple myeloma drug resistance." *Nat Commun*. 14.1 (2023): 1290.
64. Montero-Calle, Ana, et al. "In-depth quantitative proteomics analysis revealed C1GALT1 depletion in ECC-1 cells mimics an aggressive endometrial cancer phenotype observed in cancer patients with low C1GALT1 expression." *Cell Oncol*. 46.3 (2023): 697-715.
65. Kortüm, Benedikt, et al. "Combinatorial treatment with statins and niclosamide prevents CRC dissemination by unhinging the MACC1- β -catenin-S100A4 axis of metastasis." *Oncogene* 41.39 (2022): 4446-4458.
66. Nidhi, Sweta, et al. "Novel CRISPR–Cas systems: an updated review of the current achievements, applications, and future research perspectives." *Int J Mol Sci*. 22.7 (2021): 3327.

©2024, IJCSMA All Rights Reserved, www.ijcsma.com



This work is licensed under a [Creative Commons Attribution 4.0 International License](https://creativecommons.org/licenses/by/4.0/).



Espailat. A International Journal of Computer Science and Mobile Applications, Vol. 12 Issue 1, January-2024, pg. 01-16.

ISSN: 2321-8363

Impact Factor: 6.308

(An Open Accessible, Fully Refereed and Peer Reviewed Journal)

67. Chakraborty, Sajib, et al. "Onco-multi-OMICS approach: a new frontier in cancer research." *BioMed Res Int.* 2018 (2018).
68. Nurk, Sergey, et al. "The complete sequence of a human genome." *Science* 376.6588 (2022): 44-53.
69. Buniello, Annalisa, et al. "The NHGRI-EBI GWAS Catalog of published genome-wide association studies, targeted arrays and summary statistics 2019." *Nucleic Acids Res* 47.D1 (2019): D1005-D1012.
70. Sewell, Gavin W., & Arthur Kaser. "Interleukin-23 in the pathogenesis of inflammatory bowel disease and implications for therapeutic intervention." *J Crohn's Colitis* 16.Supplement_2 (2022): ii3-ii19.
71. Deelen, Joris, et al. "A meta-analysis of genome-wide association studies identifies multiple longevity genes." *Nat Commun* 10.1 (2019): 3669.
72. Wolber, Gerhard, et al. "Molecule-pharmacophore superpositioning and pattern matching in computational drug design." *Drug Discov Today* 13.1-2 (2008): 23-29.
73. Lee, Aeri, et al. "Using reverse docking for target identification and its applications for drug discovery." *Expert Opin on Drug Discov* 11.7 (2016): 707-715.
74. Nettles, James H., et al. "Bridging chemical and biological space: "target fishing" using 2D and 3D molecular descriptors." *J Med Chem* 49.23 (2006): 6802-6810.
75. McCulloch, Warren S., & Walter Pitts. "A logical calculus of the ideas immanent in nervous activity." *Bull Math Biophys* 5 (1943): 115-133.
76. Zhavoronkov, Alex, et al. "Deep learning enables rapid identification of potent DDR1 kinase inhibitors." *Nat Biotechnol* 37.9 (2019): 1038-1040.
77. Zhavoronkov, Alex, et al. "Artificial intelligence for aging and longevity research: Recent advances and perspectives." *Ageing Res Rev.* 49 (2019): 49-66.
78. Aliper, Alexander, et al. "Deep learning applications for predicting pharmacological properties of drugs and drug repurposing using transcriptomic data." *Mol Pharm* 13.7 (2016): 2524-2530.
79. Luo, Renqian, et al. "BioGPT: generative pre-trained transformer for biomedical text generation and mining." *Brief Bioinform.* 23.6 (2022): bbac409.
80. Shayakhmetov, Rim, et al. "Molecular generation for desired transcriptome changes with adversarial autoencoders." *Front Pharmacol.* 11 (2020): 509129.
81. Song, Jiali, et al. "The discovery of new drug-target interactions for breast cancer

©2024, IJCSMA All Rights Reserved, www.ijcsma.com



This work is licensed under a [Creative Commons Attribution 4.0 International License](https://creativecommons.org/licenses/by/4.0/).



Espailat. A International Journal of Computer Science and Mobile Applications, Vol. 12 Issue 1, January-2024, pg. 01-16.

ISSN: 2321-8363

Impact Factor: 6.308

(An Open Accessible, Fully Refereed and Peer Reviewed Journal)

treatment." *Molecules* 26.24 (2021): 7474.

82. Chawla, Nitesh V., et al. "SMOTE: synthetic minority over-sampling technique." *J Artif Intell Res.* 16 (2002): 321-357.
83. Achuthan, Srisairam, et al. "Leveraging deep learning algorithms for synthetic data generation to design and analyze biological networks." *J Biosci.* 47.3 (2022): 43.
84. Howe III, Edmund G., & Falcia Elenberg. "Ethical challenges posed by big data." *Innov Clin Neurosci.* 17.10-12 (2020): 24.
85. Rajotte, Jean-Francois, et al. "Synthetic data as an enabler for machine learning applications in medicine." *Iscience* 25.11 (2022).
86. Pushpakom, Sudeep, et al. "Drug repurposing: progress, challenges and recommendations." *Nat Rev Drug Discov* 18.1 (2019): 41-58.
87. Ahmed, Faheem, et al. "A comprehensive review of artificial intelligence and network based approaches to drug repurposing in Covid-19." *Biomed Pharmacother.* 153 (2022): 113350.
88. Jarada, Tamer N., et al. "A review of computational drug repositioning: strategies, approaches, opportunities, challenges, and directions." *J Cheminformatics* 12 (2020): 1-23.
89. Ashburn, Ted T., & Karl B. Thor. "Drug repositioning: identifying and developing new uses for existing drugs." *Nat rev Drug Discov* 3.8 (2004): 673-683.
90. Rodriguez, Steve, et al. "Machine learning identifies candidates for drug repurposing in Alzheimer's disease." *Nat Commun* 12.1 (2021): 1033.
91. Gupta, Rohan, et al. "Artificial intelligence to deep learning: machine intelligence approach for drug discovery." *Mol Divers* 25 (2021): 1315-1360.
92. Shen, Min, et al. "Development and validation of k-nearest-neighbor QSPR models of metabolic stability of drug candidates." *J Med Chem* 46.14 (2003): 3013-3020.
93. Susnow, Roberta G., & Steven L. Dixon. "Use of robust classification techniques for the prediction of human cytochrome P450 2D6 inhibition." *J Chem Inf Comput Sci.* 43.4 (2003): 1308-1315.
94. Suykens, Johan AK. "Support vector machines and kernel-based learning for dynamical systems modelling." *IFAC Proc Vol* 42.10 (2009): 1029-1037.
95. Le, Brian L., et al. "Transcriptomics-based drug repositioning pipeline identifies therapeutic candidates for

©2024, IJCSMA All Rights Reserved, www.ijcsma.com



This work is licensed under a [Creative Commons Attribution 4.0 International License](https://creativecommons.org/licenses/by/4.0/).



Espailat. A International Journal of Computer Science and Mobile Applications, Vol. 12 Issue 1, January-2024, pg. 01-16.

ISSN: 2321-8363

Impact Factor: 6.308

(An Open Accessible, Fully Refereed and Peer Reviewed Journal)

COVID-19." *Sci Rep.* 11.1 (2021): 12310.

96. Dönertaş, Handan Melike, et al. "Gene expression-based drug repurposing to target aging." *Aging cell* 17.5 (2018): e12819.
97. Napolitano, Francesco, et al. "Drug repositioning: a machine-learning approach through data integration." *J Cheminformatics* 5 (2013): 1-9.
98. Yang, Jihong, et al. "Drug–disease association and drug-repositioning predictions in complex diseases using causal inference–probabilistic matrix factorization." *J Chem Inf Model.* 54.9 (2014): 2562-2569.
99. Zhang, Yang, et al. "Deep learning driven drug discovery: tackling severe acute respiratory syndrome coronavirus 2." *Front Microbiol.* 12 (2021): 739684.
100. Chen, Hongming, et al. "The rise of deep learning in drug discovery." *Drug Discov Today* 23.6 (2018): 1241-1250.
101. Yang, Fan, et al. "Machine learning applications in drug repurposing." *Interdiscip Sci: Comput Life Sci.* 14.1 (2022): 15-21.

