Annual Review of Medicine

Artificial Intelligence to Guide Repurposing of Drugs

Zhimin Fu,[1] Yuxin Yang,[2,3] Mina K. Chung,[4,5,6] and Feixiong Cheng[2,3,4]

[1]Department of Pharmacy Services, University Hospitals of Cleveland, Cleveland, Ohio, USA

[2]Cleveland Clinic Genome Center, Cleveland Clinic Research, Cleveland Clinic, Cleveland, Ohio, USA; email: chengf@ccf.org

[3]Department of Genomic Sciences and Systems Biology, Cleveland Clinic Research, Cleveland Clinic, Cleveland, Ohio, USA

[4]Department of Molecular Medicine, Cleveland Clinic Lerner College of Medicine, Case Western Reserve University, Cleveland, Ohio, USA

[5]Department of Cardiovascular and Metabolic Sciences, Lerner Research Institute, Cleveland Clinic, Cleveland, Ohio, USA

[6]Department of Cardiovascular Medicine, Heart, Vascular, and Thoracic Institute, Cleveland Clinic, Cleveland, Ohio, USA

Annu. Rev. Med. 2026. 77:381–98

The *Annual Review of Medicine* is online at med.annualreviews.org

https://doi.org/10.1146/annurev-med-050224-122802

Keywords

artificial intelligence, Alzheimer's disease, drug repurposing, electronic health records, multiomics, personalized medicine

Abstract

With the pharmacokinetics, dosing, safety, and manufacturing of approved or investigational drugs already well-characterized, drug repurposing and repositioning offer emerging strategies to rapidly develop effective treatments for various challenging diseases. However, the growing mass of genetic and multiomics data has not been effectively explored by the drug repurposing community due to a lack of accurate approaches. This review aims to be an authoritative, critical, and accessible review and discussion of general interest to the drug repurposing community concerning the use of artificial intelligence (AI) and machine learning (ML) tools. Emerging questions include what is achievable with AI in this domain and what its impact will be, what AI and ML embrace, and how we, as geneticists, pharmacologists, and computational scientists, can contribute to the discovery of new, inexpensive, and affordable repurposable medicines. The fast growth of genetics and multiomics data (genomics, transcriptomics, proteomics, metabolomics, and radiomics) and electronic health records in diverse populations contributes

to answering questions, including how to rapidly identify effective repurposable medicines, what a clinically meaningful effect size in trials is, and what the potential implications for precision medicine are. This review discusses AI and ML for drug repurposing in the context of genetics, multiomics, real-world data collection, and crowdsourcing of knowledge. We conclude by considering questions on how AI and ML methodologies can unite the diverse aspects of translational medicine for emerging treatment development in human-challenging diseases.

INTRODUCTION

Although investment in biomedical and pharmaceutical research has increased significantly over the past two decades, specifically for understanding genetic risk factors of disease, we still need to develop more effective disease-modifying treatments for multiple challenging diseases, such as Alzheimer's disease (AD), heart disease, cancer, and COVID-19 (1–3). This is partly because the growing mass of genetic and multiomics datasets has not been effectively explored for drug development due to a lack of accurate approaches. Drug repurposing and repositioning reduce the time and cost of drug development (1–4). With the pharmacokinetics, dosing, safety, and manufacturing of approved or investigational drugs already well-characterized, the goal of drug repositioning is to identify new indications for drugs (4–6); for example, an anti-inflammatory agent for arthritis might be repositioned for treatment of AD or AD-related dementia (ADRD). However, how to prioritize drug targets and candidate repurposable medicines for complex human diseases at drugome-wide and genome-wide scales is challenging.

The recent increase in generation of multiomics data, including genetics, genomics, epigenomics, transcriptomics, proteomics, metabolomics, lipidomics, radiomics, and phenomics, as well as data digitalization in patient care and the pharmaceutical sector, presents both challenges and opportunities (**Figure 1**). For example, the growing availability of big data results in challenges for personalized clinical diagnosis and treatment of human disease. These challenges have motivated the application of advanced artificial intelligence (AI) and machine learning (ML) tools to help scientists identify repurposable drugs and accelerate progress toward effective treatments for a variety of challenging, complex diseases.

TARGET-CENTRIC AND DISEASE-CENTRIC DRUG REPURPOSING

There are two major ways to do drug repurposing: target-centric drug repurposing and disease-centric drug repurposing. In a recent analysis of repurposed drugs, the majority were developed by disease-centric repurposing (7). The premise of disease-centric drug repurposing is that the same drug can be applied to treat different diseases when these diseases share similar biological pathways, symptoms, or traits (7). A key step in conducting disease-centric drug repurposing is to identify underlying biological mechanisms of action for the target disease that are homologous to the original disease that a drug treats (8). Sildenafil is an oral medication used to treat pulmonary arterial hypertension and erectile dysfunction (9, 10). Sildenafil reduces the breakdown of cyclic guanosine monophosphate (cGMP) by inhibiting cGMP-specific phosphodiesterase type 5 (11). Recent work suggests sildenafil has potential for AD treatment (12). By studying AD endophenotype disease modules within protein–protein interaction (PPI) networks, along with real-world patient data showing reduced incidence of AD, sildenafil stands out as a novel candidate drug for AD (13). In vivo investigations imply that sildenafil helps bring back cognitive function in an AD mouse model by reducing activity of glycogen synthase kinase 3 and cyclin-dependent kinase 5, as well as bringing up the level of brain-derived neurotrophic factor (14). Studies using induced

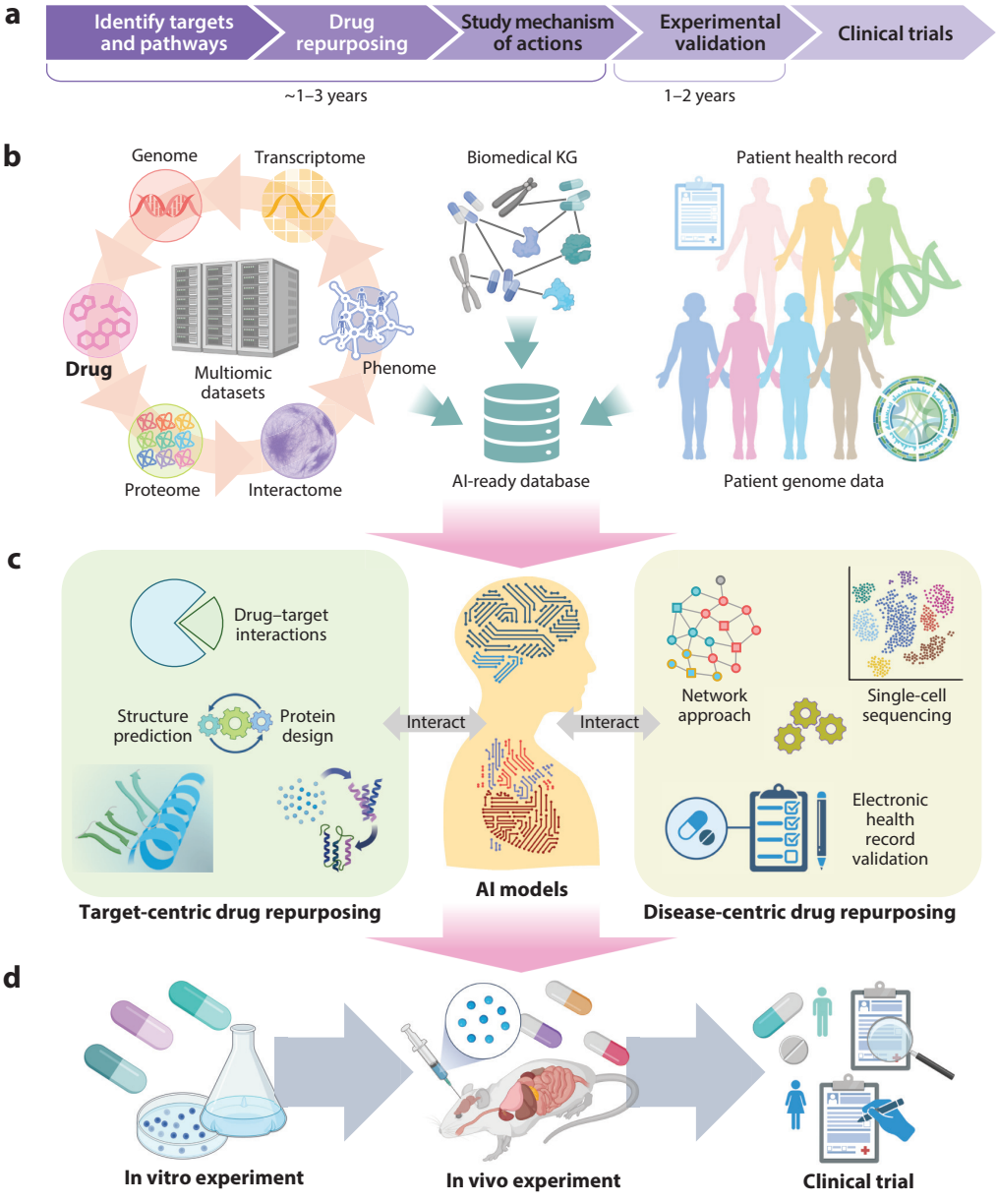

Figure 1

A proposed artificial intelligence (AI)-based drug discovery pipeline for personalized medicine. (*a*) An overview pipeline of AI-based drug repurposing. (*b*) Constructing AI-ready databases. These databases consist of three pivotal datasets: a biomedical knowledge graph (KG) dataset, multiomics dataset, and patient or clinical dataset. The biomedical KG dataset covers relations between drugs, diseases, genes, pathways, and tissues. The multiomics dataset includes data about genomes, transcriptomes, phenomes, interactomes, proteomes, and drugs, while the patient dataset contains patient health records and genomic data of patients. (*c*) AI-assisted drug approaches for both disease-centric and target-centric drug repurposing. (*d*) Experimental validation of AI-prioritized repurposable drugs. The AI-prioritized candidate drugs can be validated using in vitro and in vivo experiments to ensure their efficacy and safety. Subsequently, the candidate drugs with ideal efficacy and safety profiles will be moved to clinical trial on patients with similar diseases, symptoms, and genetic profiles under the precision medicine hypothesis. Figure created in BioRender; ZY. 2026. **https://BioRender.com/ooyt25i**.

pluripotent stem cells (iPSC) from AD patients also indicate that sildenafil targets AD-related genes and pathways (15).

On the other hand, target-centric drug repurposing assumes that the same target protein is associated with different diseases, so a drug that inhibits or activates this target protein has the potential to treat these diseases. Metformin is the first-line treatment for type 2 diabetes, working by reducing hepatic glucose production and internal absorption of glucose while improving insulin sensitivity (16). Metformin activates a cellular energy sensor, adenosine monophosphate–activated protein kinase, which is considered to be a significant pathway in atrial fibrillation (AF) through transcriptomic-based network analysis (17). Semaglutide is a glucagon-like peptide 1 (GLP-1) receptor agonist that controls type 2 diabetes, treats obesity, and reduces risk of cardiovascular diseases (18–20). Semaglutide promotes secretion of insulin from pancreatic beta cells and cuts production of glucagon from pancreatic alpha cells, which reduces fasting and postprandial plasma glucose (21). Administration of GLP-1 receptor agonists is reported to lower the rewarding effect of alcohol and drugs, thus mediating substance abuse (22). GLP-1 receptors may also be associated with neurodegenerative diseases, such as AD, via insulin signaling (23). Studies show that semaglutide demonstrates neuroprotection in a rat model via reducing inflammation and apoptosis (24). These successful examples of drug repurposing have offered effective strategies for the rapid development of potential treatments for human disease using disease-centric or target-centric approaches (**Figure 1**).

Yet, existing data, including genomics, transcriptomics, proteomics, and longitudinal real-world data (RWD), have not yet been fully utilized and integrated for disease-modifying drug repurposing for human disease. Systematic characterization and identification of underlying pathobiology could provide a foundation for identifying disease-modifying targets and repurposable medicine. Integration of the genome, transcriptome, proteome, and human interactome from diverse populations is essential for such identification using computational models, as discussed below.

ARTIFICIAL INTELLIGENCE FOR DRUG REPURPOSING

Artificial Intelligence–Ready Datasets

Effective drug discovery pipelines require multimodal biomedical data, such as genomics, transcriptomics, proteomics, metabolomics, imaging, biofluid markers, and real-world patient data (**Figure 1**), and in vivo and in vitro validation in both animal and human models. These rich databases offer necessary data for developing AI and ML models or tools for drug repurposing. There are five different types of databases (**Table 1**) that can be leveraged in drug repurposing: (*a*) chemoinformatic databases, where structures, formulas, and other properties of molecules are included; (*b*) bioinformatics databases, where structure, function, and additional information about proteins and genes are provided; (*c*) systems biology databases, where molecular reactions and interactions with diseases are stored; (*d*) multiomics databases, where details of diseases and genes and mutations associated with diseases are deposited; and (*e*) pharmacological databases, where interactions and bindings between drugs and targets are present. Commonly used chemoinformatic databases (**Table 1**) include ChEMBL (25), PubChem (26), and DrugBank (27). Protein and gene databases include UniProt (28), PDB (29), the AlphaFold Protein Structure Database (30), GenBank (31), and Ensembl (32). Systems biology and pathway databases include KEGG (33) and Reactome (34). Disease–gene databases include DISEASES (35) and DisGeNET (36). Drug–target interaction (DTI) databases include BindingDB (37) and PDBbind (38). These comprehensive databases offer rich, AI-ready datasets to build and evaluate various AI and ML models for drug repurposing and repositioning studies.

Table 1 List of AI/ML-ready databases and selected tools for drug repurposing

Database name	Database type	Description	Website
ChEMBL (25)	Chemoinformatic and/or drug–target	A database including chemical and genomic information of bioactive compounds	**https://www.ebi.ac.uk/chembl/**
DrugBank (27)		A database with drugs, drug targets, and clinical information	**https://go.drugbank.com/**
PubChem (26)		A database of molecule properties, structure, and clinical information	**https://pubchem.ncbi.nlm.nih.gov/**
BindingDB (37)		A database containing information of protein–ligand binding affinity	**https://www.bindingdb.org/**
PDBbind (38)		A database of experimentally determined binding affinity data of protein–ligand binding	**http://www.pdbbind.org.cn/**
Diseases (35)	Disease–gene	A database of disease–gene associations curated from literature, mutation data, and genome-wide association studies	**https://diseases.jensenlab.org/**
DisGeNet (123)		A database of genomic and human diseases	**https://disgenet.com/**
GeneCards (124)		A database of genomic, genetic, clinical, and functional information	**https://www.genecards.org/**
Open Targets (125)		A database of identification and prioritization of potential genes related with diseases	**https://platform.opentargets.org/**
Epic Cosmos	EHR	A database of EHR data from the Epic system	**https://cosmos.epic.com/**
TriNetX		A database of EHR data from multiple EHR systems, including claims and mortality data	**https://trinetx.com/solutions/real-world-datasets/**
AlphaFold Protein Structure Database (30)	Protein and gene	A database of AlphaFold predicted protein structures	**https://alphafold.ebi.ac.uk/**
Ensembl (126)		A database of annotated genome information	**https://www.ensembl.org/index.html**
GenBank (31)		A database of annotated genetic sequences	**https://www.ncbi.nlm.nih.gov/genbank/**
PDB (127)		A database of experimentally validated protein crystal 3D structures	**https://www.rcsb.org/**
UniProt (28)		A database of protein sequences and functions	**https://www.uniprot.org/**
Interactome INSIDER (128)	Systems biology and pathway	A protein–protein interaction database with genomic variant information	**http://interactomeinsider.yulab.org/**
KEGG (33)		A database of genes and pathways	**https://www.genome.jp/kegg/**
Reactome (129)		A database of human pathways and genes	**https://reactome.org/**
STRING (130)		A database of functional protein-protein association networks	**https://string-db.org/**

Abbreviations: AI, artificial intelligence; EHR, electronic health record; ML, machine learning.

Table 2 Commonly used data resources to develop artificial intelligence drug repurposing tools

Name	Description	Link
CHEM-BERT (53)	A molecular sequence model that predicts molecular properties and drug targets	**https://github.com/HyunSeobKim/CHEM-BERT**
MMELON (61)	A foundation multimodal model that integrates different modalities of small molecules to predict molecular properties and drug targets	**https://github.com/BiomedSciAI/biomed-multi-view**
ImageMol (57)	A foundation model that trained on molecule images to predict drug targets and molecule properties	**https://github.com/ChengF-Lab/ImageMol**
VideoMol (58)	A foundation model that converts molecule images to videos to study molecule properties and activities	**https://github.com/ChengF-Lab/VideoMol**
deepDR (131)	A network-based model combining 10 types of networks to repurpose drugs	**https://github.com/ChengF-Lab/deepDR**
deepDTnet (66)	A network-based model combining 15 types of networks to identify drug targets	**https://github.com/ChengF-Lab/deepDTnet**
LISA-CPI (132)	A molecular image model combined with target 3D structure information to predict drug target binding activity	**https://github.com/ChengF-Lab/LISA-CPI**
Chemception (56)	A molecular image model that predicts activity on drug targets	**https://github.com/Abdulk084/Chemception**
MolCLR (133)	A molecular structure-based model that can predict molecular properties and drug targets	**https://github.com/yuyangw/MolCLR**

Machine Learning and Deep Learning Techniques

ML and deep learning (DL) are subfields of AI that utilize different algorithms to learn patterns from input data (39). ML and DL algorithms can be categorized into three major types: (*a*) supervised learning, where label information is required; (*b*) unsupervised learning, where no label information is needed; and (*c*) semisupervised learning, where labels for only a small portion of the data are necessary (**Table 2**). ML algorithms usually use structured or vector data. Molecular fingerprints, which capture chemical features of molecules, are commonly used as input in ML algorithms. Circular fingerprints (40), molecular access system fingerprints (41), and PubChem fingerprints (42) are widely adopted to represent structures and chemical properties of molecules. DL methods can be applied to unstructured data, such as simplified molecular input line entry system (SMILES) (43) formulas, amino acid sequences, molecular images, 3D structures, and electronic health records (EHRs). DTIs are primarily used in target-centric drug repurposing (**Figure 1**). Several early studies combined fingerprints and ML models to predict DTIs, which is a fundamental step in target-centric drug repurposing (44–46). Beyond using molecular fingerprints, similarities between drugs or targets can also be used as inputs to classical ML algorithms. In this case, similarities between chemical structure, side effect profile, amino acid sequence, and gene expression responses of drugs and targets can be measured using different metrics, followed by generating similarity matrices to recognize potential repurposable drugs and their corresponding targets for diseases (47). Jacob & Vert (48), Bleakley et al. (49), and Mordelet & Vert (50) developed several models based on classic ML techniques to leverage similarity matrices for target-centric drug repurposing.

Sequence models are user-friendly as they do not require difficult work to obtain fingerprints or similarity matrices. Smiles2vec (51), DeepSMILES (52), CHEM-BERT (53), and SMILES-BERT (54) are several notable examples that leverage sequence data (**Table 2**). Sequence models can achieve promising results, yet they lack information about the 3D structure of molecules and proteins, which is essential to the function of drugs and targets. Molecular image data provide

2D information about molecular structure, leading to better performance. DEEPScreen (55) and Chemception (56) are early examples that used molecular images in drug discovery. ImageMol exploits the power of pretraining to improve accuracy and generalizability (57). There are two primary ways to encode 3D structures: using nodes and edges in geometric data to represent atoms and connections and saving molecules from different angles as a video. A notable example of the latter is VideoMol (58), where each molecule is encoded in a video to elucidate the conformational changes of molecules. Another trend in DL techniques is multimodality, where distinct modalities of data, such as sequence data and structure data, are fused into one DL model to provide comprehensive understanding of drugs and targets (59). MRL-Mol (60) and MMELON (61) provide pilot studies in combining diverse modalities of molecules, while CLEAN-Contact (62) achieves better performance by incorporating different modalities of proteins.

Network-Based Approaches

Network-based techniques are largely used in disease-centric drug repurposing, as they use information derived from disease-related networks to find repurposable drugs for a specific disease (63). Network-based techniques use graph algorithms to analyze graph data, which consist of nodes and edges. In particular, nodes can represent drugs, targets, patients, genes, or pathways, while edges represent relationships between different entities. Widely used networks include the PPI network, which helps with understanding how protein functions within cells, and the drug–drug–disease network, which offers insight into how different drugs can interact with each other and cure diseases. Network-based techniques stand out at learning undiscovered relations between drugs and their potential targets, as well as between diseases and their target proteins. A systems pharmacology–based network approach was developed and identified hydroxychloroquine, an antimalarial and antirheumatic drug, as a potential repurposable drug to reduce risk of cardiovascular diseases (CVDs) (64). More recently, a genome-wide positioning systems network (GPSnet) (65) was developed to aim at human PPI networks with patients' DNA and RNA sequences mapped. Ouabain, a cardiac arrhythmia and heart failure drug, was identified with antitumor activity in lung adenocarcinoma. DeepDTnet (66) achieves high accuracy by integrating 15 different types of networks and predicts topotecan, a topoisomerase inhibitor, to be a potential therapy against multiple sclerosis by inhibiting human retinoic acid receptor–related orphan receptor-gamma t (ROR-γt). By concentrating on an endophenotype, an intermediate characteristic on the biological pathways between genotype and disorder, an in silico network medicine approach found sildenafil as a potential treatment for AD (13).

Clinical Trial Emulation from Real-World Patient Databases

Randomized controlled trials (RCTs) are the gold standard for drug development (67). However, due to the stringent definition of eligibility criteria, the number of eligible trial participants is typically small. This makes the trial cohort ideal rather than representative of real-world patient populations. RWD are comprised of practice-based observations from real-world patients and thus provide a valuable resource for population-based validation of drug repurposing (13, 68). However, RWD is challenged by confounding factors, such as sex, race, and socioeconomic status, as well as a lack of detailed clinical, biomarker, and genetic information, as well as other unknown factors. Recently, there have been attempts to replicate the treatment effects obtained from RCTs in RWD with the help of AI techniques. RWD cohorts with features like those of the trial participants are identified, and if individual data are available for RCT participants then (weighted) propensity score matching approaches can be applied. Due to complicated confounding factors in RWD (e.g., high dimensionality and temporality), classical ML approaches are unable to estimate the

propensity scores with high accuracy. DL approaches can resolve this problem through the target trial emulation method (69, 70). This method can be applied in large-scale insurance claims, and it found zolpidem [a US Food and Drug Administration (FDA)-approved anti-insomnia medicine] as a potential drug for slowing progression of Parkinson's dementia (71). Using a DL framework on emulating clinical trials from RWD (72), researchers identified 14 drug candidates that reduced risk of AD in patient subgroups with specific clinical features. Under high-throughput target trial emulation, the team identified five top-ranked drugs (pantoprazole, gabapentin, atorvastatin, fluticasone, and omeprazole) originally intended for other indications with potential benefits for AD patients (73).

Due to the complex confounding situations in RWD, conventional propensity score estimation approaches based on logistic regression may not be able to accurately estimate the probability of treatment and censoring. Advanced AI approaches may achieve this goal. Recent studies have evaluated three types of strategies for the representation of EHR data for predictions: (*a*) convolutional neural network–based methods, which represent the EHRs of each patient as an event in a time matrix and perform a series of one-side convolutions, activations plus pooling, followed by a final, fully connected layer to perform the predictions (74); (*b*) recurrent neural network (RNN)-based methods, which represents each individual's EHR as an event sequence and can be used to train an RNN-based prediction model (75); and (*c*) graph-based methods, which represent the EHR as a heterogeneous information network with patients and clinical events as nodes and event co-occurrences as edges, applying graph neural network–based approaches to perform the predictions (76). Yet, a recent study showed that the DL-based propensity score model did not necessarily outperform logistic regression–based methods in confounding factor justification (73). In addition, data missingness is another feature of noisy RWD. To handle missingness not at random, researchers can consider selection through model-based methods [e.g., outcome-dependent sampling in longitudinal outcomes (77)]. In addition, one can also utilize DL-based imputation methods such as the last observation carry-forward (78) and generative-adversarial nets (79).

Biomedical Knowledge Graph for Drug Repurposing Hypothesis Generation

With the rich, relevant, and high-quality knowledge found in the literature, one can both validate the drug repurposing hypotheses generated from RWD and generate new knowledge-based repurposing hypotheses. A recent study first generated embeddings for the entities (drugs, diseases, or genes/proteins) and relations (i.e., drug–disease associations) using knowledge graph embedding methods such as Deep Graph Library–Knowledge Embedding (DGL-KE) (80), suitable for large-scale data. These embedding methods can generate vector-based representations for both entities and relations in the same embedded semantic space, such that the repurposing hypotheses can be generated according to the ranking scores of <*drug*, *relation*, *disease*> triples evaluated with vector-based similarities. Using a graph foundation model for zero-shot drug repurposing from a large biomedical knowledge graph, a graph neural network and metric learning module with high accuracy were developed to rank drugs as potential indications and contraindications for 17,080 diseases, including a large number of rare diseases (81).

Other research developed a comprehensive biomedical knowledge graph concerning COVID-19 (82). This COVID-19 knowledge graph consists of 15 million edges and 39 different types of relations, including connections between drugs, diseases, genes, anatomies, pharmacologic classes, and gene expression. Similar to previous research, an embedding model was developed to extract representation vectors for <*head entity*, *relation type*, *tail entity*>. A DL model trained on this knowledge graph identified more than 40 potential repurposable drugs for COVID-19. Additional enrichment analysis was conducted to validate the predicted repurposable drugs that

show high confidence in treating COVID-19. Such biomedical knowledge graphs accompanied by highly accurate DL models are important AI methods for advancing drug repurposing.

Clinical and Experimental Validation of Artificial Intelligence–Based Drug Repurposing

Experimental and clinical validation are crucial steps after repurposing drugs using AI models, as such validations could confirm their real-world accuracy and reliability. Primary approaches to conduct experimental validation include using iPSC-derived models and animal models (68), while clinical validation may use EHR or health insurance claim data (**Figure 1**). In recent work where sildenafil was repurposed as a potential AD treatment, iPSC-derived neurons from AD patients treated with sildenafil showed significant reduction in phosphorylated-tau 181 (p-tau 181), an early biomarker of AD pathology (13). Other researchers conducted additional experiments to validate the effectiveness of sildenafil on the reduction of both p-tau 181 and p-tau 205 (15). Additionally, enrichment analysis of differentially expressed genes between control and sildenafil-treated groups displayed a neuroprotective effect. Another work that repurposed metformin as a therapeutic treatment for AF leveraged human iPSC-derived atrial-like cardiomyocytes to validate in silico predictions (17). Critical cardiovascular-related markers were significantly upregulated, which was accompanied by an increase in expression of several known markers via metformin associated with low expression in AF. In another study, iPSC-derived microglia from AD patients were treated with ketorolac, a moderate-to-severe pain treatment, by downregulating the type-I interferon signaling, mechanistically supporting the potential benefits of ketorolac in reducing the incidence of AD by targeting disease-relevant microglia (83).

APPLICATIONS OF ARTIFICIAL INTELLIGENCE APPROACHES TO DRUG REPURPOSING

Critically, drug repurposing depends on efficient searching of the vast drug space, for which the optimal approach is rapidly evolving. As drug repurposing is a complex process involving many steps, multimodal ML tools can significantly reduce the time and cost of drug development. For instance, multimodal ML approaches (84, 85) improve the accuracy of patient subphenotyping during clinical trial design by assembling neuroimaging, genetic, and multiomics profiling data. With the help of DL, effective representations can be learned for different data modalities (86, 87), which can then be fused by simple concatenation or more complicated nonlinear transformation (88) to perform downstream tasks such as molecular design, pharmacokinetics property evaluation, and optimization. AI and ML tools have become a leading technique for expediting and reducing the cost of the drug development. We now turn to using four types of challenging human complex diseases (including AD, cancer, COVID-19, and CVDs) to illustrate how AI accelerates the finding of treatments by repurposing approved drugs.

There are 50 repurposable drug trials (40 unique repurposed agents) based on the latest AD drug development pipeline in 2024 (89). To expand drug repurposing efforts, a recent ML-based framework named DRIAD (drug repurposing in AD) has been developed to quantify the potential associations between AD biological processes and linked genetic datasets, thereby prioritizing drug candidates for repurposing (90). DRIAD prioritized baricitinib as a candidate AD drug, and baricitinib is being tested in an open-label, biomarker-driven basket trial (NCT05189106) in people with AD and amyotrophic lateral sclerosis.

Another recently developed tool, AlzGPS, is a systems biology platform with over 100 multiomics datasets that capture molecular profiles underlying AD pathobiology (91). This tool enables network-based prioritization of potential targets for AD drug repurposing

(91). NETTAG is a network topology–based DL framework to identify disease-associated genes for AD and prioritize candidate drugs and targets. Using NETTAG (92), the team successfully identified gemfibrozil (an approved lipid regulator) as being significantly associated with reduced risk of AD compared with simvastatin using an active-comparator design from real-world patient data. A DL methodology (deepDTnet) was developed for new target identification and drug repurposing in a heterogeneous drug–gene–disease network embedded with 15 types of chemical, genomic, phenotypic, and cellular network profiles (66) (**Table 1**). Trained on 732 US FDA-approved small molecule drugs, deepDTnet shows high accuracy in identifying novel molecular targets for known drugs, outperforming previously published state-of-the-art methodologies. Importantly, the authors experimentally validated deepDTnet-predicted topotecan as a candidate repurposable drug for multiple sclerosis by targeting human ROR-γt (66).

Using multimodal analysis of single-cell or nucleus RNA-sequencing data from AD patient brains, two approved asthma drugs (fluticasone and mometasone) were found to be significantly associated with reduced likelihood of AD by targeting AD-associated microglia (93). Via analysis of real-world electronic insurance record data from 7.2 million patients from the MarketScan Medicare Supplemental Database, two FDA-approved p300/CBP inhibitors, salsalate and diflunisal, were found to be associated with decreased incidence of AD, and neuroprotective efficacy was also validated in mice (94, 95). Using an endophenotype-based in silico network medicine approach (13), one team showed that sildenafil usage was significantly associated with reduced likelihood of AD and further validated the findings using an AD patient iPSC-derived neuron model (13). Another team demonstrated that bumetanide (an FDA-approved oral diuretic) provides a potential treatment for apolipoprotein *APOE4*-related AD using in silico approaches combining experimental and real-world evidence (96). These prototypical examples illustrate how AI-based multiomics approaches combined with real-world patient databases and experimental approaches can rapidly identify potential treatments for AD (**Figure 1**). The details of computational methods for AD drug repurposing can be found in other recent reviews (4–6).

Cancers are challenging diseases that are caused by malignant tumors involving uncontrolled growth of cells in the body (97). Cancer is the second biggest contributor to death in the United States, with lung cancer as the deadliest cancer (98). A major method to apply AI to repurpose drugs for cancer is leveraging cancer cell line models and patient-derived primary cells (99). An ML algorithm based on drug response to a human breast cancer cell line found 16 out of 28 drugs had significantly better performance in treating human breast cancer. Additionally, EHRs of cancer patients can be used for predicting drug effectiveness. For example, investigations have been made to discover potential drugs that can lower cancer mortality by using EHRs from diverse healthcare systems and/or medical databases (100).

COVID-19 is an infectious disease caused by the SARS-CoV-2 virus, which became a global pandemic (101). In the early stage of the pandemic, AI was a key technique for discovering drugs repurposable to the treatment of COVID-19 (102). Studies were focused on inhibiting SARS-CoV-2 spike proteins, key to the infection of the virus, and reducing the inflammation and immune response caused by the disease, which play a vital role in the mortality of patients. A network-based approach identified 16 drugs and 3 drug combinations with the potential to treat COVID-19 (103). ImageMol has been used to find inhibitors of SARS-CoV-2 spike proteins (57). Remarkably, melatonin, a hormone that regulates sleep, and dexamethasone, a glucocorticoid agonist, were proposed to reduce inflammation and immune response, and toremifene was presented as a potential SARS-CoV-2 spike protein inhibitor (82, 104, 105).

CVD refers to a group of diseases and disorders concerning heart and blood vessels (106). CVD is the primary cause of death around the globe. Main symptoms of CVD include chest pain, palpitations, and shortness of breath. From studying the human protein–protein interactome,

carbamazepine, a drug to control seizures, is believed to increase the risk of CVD, while hydroxychloroquine, a malaria treatment, is considered to lower the risk of CVD (64). By incorporating information about adverse effects of drugs, approved drugs have been repurposed for CVD treatment while lowering potential adverse effects (107).

DISCUSSION, PERSPECTIVE, AND FUTURE DIRECTIONS

In this review, we briefly summarize and discuss existing AI techniques and their applications to repurposing drugs for several challenging diseases. Despite substantial recent advancements in AI, the research circles of AI and biomedicine still face significant challenges. One formidable challenge is that the selection of hyperparameters, such as the learning rate and size of hidden layers, can substantially affect a model's performance. Early strategies to find optimal hyperparameters mainly used random search and grid search, which can be computationally expensive for complex AI models with large hyperparameter space (108). Recent development of automated machine learning that incorporates neural architecture search provides a viable solution for this issue (109). Even though an accurate AI model is trained with high-quality, labeled datasets, experimental validation of drug discovery results is still required for confirmation.

Recent growth in large language models (LLMs) provides heartening solutions for advancing drug repurposing through the integration of heterogeneous data sources. These AI models can have hundreds of billions to trillions of parameters while trained on innumerable data and documents (110). Such training equips LLMs with the ability to find relations between drugs, diseases, targets, and pathways. Yet LLMs trained on generic data are not immune to hallucinations and biased responses (111). Retrieval-augmented generation is a method where a specialized dataset is used to supplement LLMs to reduce hallucinations and biased responses. Moreover, AI agents, emerging AI systems designed to automatically carry out tasks and complete goals with limited human interaction, are another promising direction for drug discovery (112, 113). A future AI agentic system based on LLMs could execute different kinds of drug discovery inquiries by leveraging multiple state-of-the-art drug repurposing tools, such as AlphaFold3 (114) and NETTAG (92) (**Figure 1**). By integrating biomedical knowledge graphs into the LLMs of AI agents, patients, physicians, and researchers can ask any questions regarding drugs and diseases. For example, physicians can ask the AI agent to find a repurposable drug for a patient, as the patient may display severe adverse effects from existing drugs for the disease.

AI-based drug repurposing communities are leading big data and open science. Data used for training AI models must be annotated by domain experts to ensure the quality of AI models before effectively advancing the drug development process for human complex diseases. The application of AI solutions to drug repurposing is possible via cross-disciplinary teamwork and cooperation to address current gaps and challenges, such as analysis of clinical trial data and application of trial outcomes to patients who could benefit greatly from repurposable drugs. Altogether, AI is an indispensable aspect of the future of drug repurposing and precision medicine for challenging human diseases.

There are several intimidating challenges in building current AI-ready datasets. First, multiomics and clinical data are generated from heterogeneous patient samples across different laboratories and health care systems. Harmonization of clinical and multiomics data plays a crucial role in securing the quality of AI models in real-world drug repurposing, and data harmonization is challenging for most basic and clinical scientists. Limited data sharing is another hurdle in AI-based drug repurposing. Biopharmaceutical companies have generated massive data compared with academic institutions, and they cannot be shared due to various intellectual property issues. Because of the complexity of human diseases, heterogeneous datasets covering genomic, cellular,

clinical, and behavioral aspects are required to advance disease understanding (115, 116). Collaborations across different entities and institutions therefore need to incorporate diverse patient populations, as well as participants with diverse disease characteristics (117), using AI technologies.

Data security is another key to inclusiveness and maintenance of trust in AI technologies. For example, fast healthcare interoperability resources (118) for clinical data transmission are crucial to data credibility and confidentiality. Another potential solution is federated learning (119), which protects individual patient data through collective learning from multiple local sites without transferring the original raw data. Another important direction is to improve transparency and interpretability of AI models, such as with explainable AI technologies, so that scientists can understand the decision-making process during drug repurposing in order to ensure proper assessments of usability and potential failure cases (92). Thus, model interpretability can be contextualized (120), with different levels of model transparency required for different applications. Model interpretation methods have various potential uses (121). For example, knowledge distillation is a popular interpretation technique that aims to advance a secondary interpretable model to approximate the prediction results of the first model (122). These models could be vulnerable to adversarial attacks, which could manipulate the model explanations deliberately and make them unreliable. Deriving robust, reliable, and secure model interpretations is therefore an important future research direction.

DISCLOSURE STATEMENT

M.K.C. has received research grants from the American Heart Association. F.C. has received research grants from the Alzheimer's Drug Discovery Foundation and the Alzheimer's Association.

ACKNOWLEDGMENTS

This work was primarily supported by the National Institute on Aging under awards U01AG073323, R01AG092591, R33AG083003, R01AG066707, R01AG076448, R01AG082118, RF1AG082211, RF1NS133812, R01AG092462, R01AG084250, 1OT2OD038083-01, and U01CA294154 to F.C. and by the National Heart, Lung, and Blood Institute under award P01HL158502 to M.K.C.

LITERATURE CITED

1. Zhou Y, Wang F, Tang J, et al. 2020. Artificial intelligence in COVID-19 drug repurposing. *Lancet Digit. Health* 2:e667–76. **https://doi.org/10.1016/S2589-7500(20)30192-8**
2. Zhou Y, Hou Y, Shen J, et al. 2020. Network-based drug repurposing for novel coronavirus 2019-nCoV/SARS-CoV-2. *Cell Discov.* 6:14. **https://doi.org/10.1038/s41421-020-0153-3**
3. Zhou Y, Liu Y, Gupta S, et al. 2023. A comprehensive SARS-CoV-2-human protein-protein interactome reveals COVID-19 pathobiology and potential host therapeutic targets. *Nat. Biotechnol.* 41:128–39. **https://doi.org/10.1038/s41587-022-01474-0**
4. Fang J, Pieper AA, Nussinov R, et al. 2020. Harnessing endophenotypes and network medicine for Alzheimer's drug repurposing. *Med. Res. Rev.* 40:2386–426. **https://doi.org/10.1002/med.21709**
5. Paranjpe MD, Taubes A, Sirota M. 2019. Insights into computational drug repurposing for neurodegenerative disease. *Trends Pharmacol. Sci.* 40:565–76. **https://doi.org/10.1016/j.tips.2019.06.003**
6. Ballard C, Aarsland D, Cummings J, et al. 2020. Drug repositioning and repurposing for Alzheimer disease. *Nat. Rev. Neurol.* 16:661–73. **https://doi.org/10.1038/s41582-020-0397-4**
7. Parisi D, Adasme MF, Sveshnikova A, et al. 2020. Drug repositioning or target repositioning: a structural perspective of drug-target-indication relationship for available repurposed drugs. *Comput. Struct. Biotechnol. J.* 18:1043–55. **https://doi.org/10.1016/j.csbj.2020.04.004**

8. Boller F, Mizutani T, Roessmann U, Gambetti P. 1980. Parkinson disease, dementia, and Alzheimer disease: clinicopathological correlations. *Ann. Neurol.* 7:329–35. **https://doi.org/10.1002/ana.410070408**
9. Galiè N, Ghofrani HA, Torbicki A, et al. 2005. Sildenafil citrate therapy for pulmonary arterial hypertension. *New Engl. J. Med.* 353:2148–57. **https://doi.org/10.1056/NEJMoa050010**. Correction. 2006. *New Engl. J. Med.* 354:2400–1
10. Goldstein I, Lue TF, Padma-Nathan H, et al. 1998. Oral sildenafil in the treatment of erectile dysfunction. *New Engl. J. Med.* 338:1397–404. **https://doi.org/10.1056/NEJM199805143382001**. Correction. 1998. *New Engl. J. Med.* 339:59
11. Andersson KE. 2018. PDE5 inhibitors – pharmacology and clinical applications 20 years after sildenafil discovery. *Br. J. Pharmacol.* 175:2554–65. **https://doi.org/10.1111/bph.14205**
12. Sanders O. 2020. Sildenafil for the treatment of Alzheimer's disease: a systematic review. *J. Alzheimers Dis. Rep.* 4:91–106. **https://doi.org/10.3233/ADR-200166**
13. Fang J, Zhang P, Zhou Y, et al. 2021. Endophenotype-based in silico network medicine discovery combined with insurance record data mining identifies sildenafil as a candidate drug for Alzheimer's disease. *Nat. Aging* 1:1175–88. **https://doi.org/10.1038/s43587-021-00138-z**
14. Cuadrado-Tejedor M, Hervias I, Ricobaraza A, et al. 2011. Sildenafil restores cognitive function without affecting β-amyloid burden in a mouse model of Alzheimer's disease. *Br. J. Pharmacol.* 164:2029–41. **https://doi.org/10.1111/j.1476-5381.2011.01517.x**
15. Gohel D, Zhang P, Gupta AK, et al. 2024. Sildenafil as a candidate drug for Alzheimer's disease: real-world patient data observation and mechanistic observations from patient-induced pluripotent stem cell-derived neurons. *J. Alzheimers Dis.* 98:643–57. **https://doi.org/10.3233/JAD-231391**
16. Rena G, Hardie DG, Pearson ER. 2017. The mechanisms of action of metformin. *Diabetologia* 60:1577–85. **https://doi.org/10.1007/s00125-017-4342-z**
17. Lal JC, Mao C, Zhou Y, et al. 2022. Transcriptomics-based network medicine approach identifies metformin as a repurposable drug for atrial fibrillation. *Cell Rep. Med.* 3:100749. **https://doi.org/10.1016/j.xcrm.2022.100749**
18. Bergmann NC, Davies MJ, Lingvay I, Knop FK. 2023. Semaglutide for the treatment of overweight and obesity: a review. *Diabetes Obes. Metab.* 25:18–35. **https://doi.org/10.1111/dom.14863**
19. Andersen A, Knop FK, Vilsbøll T. 2021. A pharmacological and clinical overview of oral semaglutide for the treatment of type 2 diabetes. *Drugs* 81:1003–30. **https://doi.org/10.1007/s40265-021-01499-w**
20. Marso SP, Bain SC, Consoli A, et al. 2016. Semaglutide and cardiovascular outcomes in patients with type 2 diabetes. *New Engl. J. Med.* 375:1834–44. **https://doi.org/10.1056/NEJMoa1607141**
21. Shah M, Vella A. 2014. Effects of GLP-1 on appetite and weight. *Rev. Endocr. Metab. Disorders* 15:181–87. **https://doi.org/10.1007/s11154-014-9289-5**
22. Shirazi RH, Dickson SL, Skibicka KP. 2013. Gut peptide GLP-1 and its analogue, exendin-4, decrease alcohol intake and reward. *PLOS ONE* 8:e61965. **https://doi.org/10.1371/journal.pone.0061965**
23. Hölscher C. 2018. Novel dual GLP-1/GIP receptor agonists show neuroprotective effects in Alzheimer's and Parkinson's disease models. *Neuropharmacology* 136:251–59. **https://doi.org/10.1016/j.neuropharm.2018.01.040**
24. Yang X, Feng P, Zhang X, et al. 2019. The diabetes drug semaglutide reduces infarct size, inflammation, and apoptosis, and normalizes neurogenesis in a rat model of stroke. *Neuropharmacology* 158:107748. **https://doi.org/10.1016/j.neuropharm.2019.107748**
25. Gaulton A, Bellis LJ, Bento AP, et al. 2012. ChEMBL: a large-scale bioactivity database for drug discovery. *Nucleic Acids Res.* 40:D1100–7. **https://doi.org/10.1093/nar/gkr777**
26. Kim S, Chen J, Cheng T, et al. 2023. PubChem 2023 update. *Nucleic Acids Res.* 51:D1373–80. **https://doi.org/10.1093/nar/gkac956**
27. Knox C, Wilson M, Klinger CM, et al. 2024. DrugBank 6.0: the DrugBank Knowledgebase for 2024. *Nucleic Acids Res.* 52:D1265–75. **https://doi.org/10.1093/nar/gkad976**
28. The UniProt Consortium. 2019. UniProt: a worldwide hub of protein knowledge. *Nucleic Acids Res.* 47:D506–15. **https://doi.org/10.1093/nar/gky1049**
29. Berman HM, Westbrook J, Feng Z, et al. 2000. The Protein Data Bank. *Nucleic Acids Res.* 28:235–42. **https://doi.org/10.1093/nar/28.1.235**

30. Varadi M, Anyango S, Deshpande M, et al. 2022. AlphaFold Protein Structure Database: massively expanding the structural coverage of protein-sequence space with high-accuracy models. *Nucleic Acids Res.* 50:D439–44. **https://doi.org/10.1093/nar/gkab1061**
31. Benson DA, Cavanaugh M, Clark K, et al. 2013. GenBank. *Nucleic Acids Res.* 41:D36–42. **https://doi.org/10.1093/nar/gks1195**
32. Cunningham F, Allen JE, Allen J, et al. 2022. Ensembl 2022. *Nucleic Acids Res.* 50:D988–95. **https://doi.org/10.1093/nar/gkab1049**
33. Kanehisa M, Goto S. 2000. KEGG: Kyoto Encyclopedia of Genes and Genomes. *Nucleic Acids Res.* 28:27–30. **https://doi.org/10.1093/nar/28.1.27**
34. Gillespie M, Jassal B, Stephan R, et al. 2022. The Reactome Pathway Knowledgebase 2022. *Nucleic Acids Res.* 50:D687–92. **https://doi.org/10.1093/nar/gkab1028**
35. Grissa D, Junge A, Oprea TI, Jensen LJ. 2022. Diseases 2.0: a weekly updated database of disease–gene associations from text mining and data integration. *Database* 2022:baac019. **https://doi.org/10.1093/database/baac019**
36. Piñero J, Ramírez-Anguita JM, Saüch-Pitarch J, et al. 2020. The DisGeNET knowledge platform for disease genomics: 2019 update. *Nucleic Acids Res.* 48:D845–55. **https://doi.org/10.1093/nar/gkz1021**
37. Liu T, Lin Y, Wen X, et al. 2007. BindingDB: a web-accessible database of experimentally determined protein–ligand binding affinities. *Nucleic Acids Res.* 35:D198–201. **https://doi.org/10.1093/nar/gkl999**
38. Wang R, Fang X, Lu Y, et al. 2005. The PDBbind Database: methodologies and updates. *J. Med. Chem.* 48:4111–19. **https://doi.org/10.1021/jm048957q**
39. Jordan MI, Mitchell TM. 2015. Machine learning: trends, perspectives, and prospects. *Science* 349:255–60. **https://doi.org/10.1126/science.aaa8415**
40. Glem RC, Bender A, Arnby CH, et al. 2006. Circular fingerprints: flexible molecular descriptors with applications from physical chemistry to ADME. *IDrugs* 9:199–204
41. Durant JL, Leland BA, Henry DR, Nourse JG. 2002. Reoptimization of MDL keys for use in drug discovery. *J. Chem. Inform. Comput. Sci.* 42:1273–80. **https://doi.org/10.1021/ci010132r**
42. Fernández-de Gortari E, García-Jacas CR, Martinez-Mayorga K, Medina-Franco JL. 2017. Database fingerprint (DFP): an approach to represent molecular databases. *J. Cheminform.* 9:9. **https://doi.org/10.1186/s13321-017-0195-1**
43. Weininger D. 1988. SMILES, a chemical language and information system. 1. Introduction to methodology and encoding rules. *J. Chem. Inform. Comput. Sci.* 28:31–36. **https://doi.org/10.1021/ci00057a005**
44. Wang Y-C, Yang Z-X, Wang Y, Deng N-Y. 2010. Computationally probing drug-protein interactions via support vector machine. *Lett. Drug Design Discov*. 7:370–78. **https://doi.org/10.2174/157018010791163433**
45. Ahn S, Lee SE, Kim M-H. 2022. Random-forest model for drug–target interaction prediction via Kullback–Leibler divergence. *J. Cheminform.* 14:67. **https://doi.org/10.1186/s13321-022-00644-1**
46. Shi H, Liu S, Chen J, et al. 2019. Predicting drug-target interactions using Lasso with random forest based on evolutionary information and chemical structure. *Genomics* 111:1839–52. **https://doi.org/10.1016/j.ygeno.2018.12.007**
47. Mitchell JBO. 2001. The relationship between the sequence identities of alpha helical proteins in the PDB and the molecular similarities of their ligands. *J. Chem. Inform. Comput. Sci.* 41:1617–22. **https://doi.org/10.1021/ci010364q**
48. Jacob L, Vert J-P. 2008. Protein-ligand interaction prediction: an improved chemogenomics approach. *Bioinformatics* 24:2149–56. **https://doi.org/10.1093/bioinformatics/btn409**
49. Bleakley K, Biau G, Vert J-P. 2007. Supervised reconstruction of biological networks with local models. *Bioinformatics* 23:i57–65. **https://doi.org/10.1093/bioinformatics/btm204**
50. Mordelet F, Vert J-P. 2008. SIRENE: supervised inference of regulatory networks. *Bioinformatics* 24:i76–82. **https://doi.org/10.1093/bioinformatics/btn273**
51. Goh GB, Hodas NO, Siegel C, Vishnu A. 2017. Smiles2vec: an interpretable general-purpose deep neural network for predicting chemical properties. Preprint, arXiv:1712.02034 [stat.ML]
52. Berenger F, Tsuda K. 2021. Molecular generation by Fast Assembly of (Deep)SMILES fragments. *J. Chemoinform.* 13:88. **https://doi.org/10.1186/s13321-021-00566-4**

53. Kim H, Lee J, Ahn S, Lee JR. 2021. A merged molecular representation learning for molecular properties prediction with a web-based service. *Sci. Rep.* 11:11028. **https://doi.org/10.1038/s41598-021-90259-7**
54. Wang S, Guo Y, Wang Y, et al. 2019. SMILES-BERT: large scale unsupervised pre-training for molecular property prediction. In *Proceedings of the 10th ACM International Conference on Bioinformatics, Computational Biology and Health Informatics*. Association for Computing Machinery. **https://doi.org/10.1145/3307339.3342186**
55. Rifaioglu AS, Nalbat E, Atalay V, et al. 2020. DEEPScreen: high performance drug–target interaction prediction with convolutional neural networks using 2-D structural compound representations. *Chem. Sci.* 11:2531–57. **https://doi.org/10.1039/C9SC03414E**
56. Goh GB, Siegel C, Vishnu A, et al. 2017. Chemception: a deep neural network with minimal chemistry knowledge matches the performance of expert-developed QSAR/QSPR models. Preprint, arXiv:1706.06689 [stat.ML]
57. Zeng X, Xiang H, Yu L, et al. 2022. Accurate prediction of molecular properties and drug targets using a self-supervised image representation learning framework. *Nat. Mach. Intell.* 4:1004–16. **https://doi.org/10.1038/s42256-022-00557-6**
58. Xiang H, Zeng L, Hou L, et al. 2024. A molecular video-derived foundation model for scientific drug discovery. *Nat. Commun.* 15:9696. **https://doi.org/10.1038/s41467-024-53742-z**
59. Ngiam J, Khosla A, Kim M, et al. 2011. Multimodal deep learning. In *Proceedings of the 28th International Conference on Machine Learning*, ed. L Getoor, T Scheffer. Omnipress
60. Yang Y, Wang Z, Ahadian P, et al. 2024. A deep multimodal representation learning framework for accurate molecular properties prediction. In *Proceedings of the Great Lakes Symposium on VLSI 2024*, ed. I Partin-Vaisband, S Katkoori, L Peng, et al. Association for Computing Machinery. **https://doi.org/10.1145/3649476.3660377**
61. Suryanarayanan P, Qiu Y, Sethi S, et al. 2024. Multi-view biomedical foundation models for molecule-target and property prediction. Preprint, arXiv:2410.19704 [q-bio.BM]
62. Yang Y, Jerger A, Feng S, et al. 2024. Improved enzyme functional annotation prediction using contrastive learning with structural inference. *Commun. Biol.* 7:1690. **https://doi.org/10.1038/s42003-024-07359-z**
63. Cheng F, Desai RJ, Handy DE, et al. 2018. Network-based approach to prediction and population-based validation of in silico drug repurposing. *Nat. Commun.* 9:2691. **https://doi.org/10.1038/s41467-018-05116-5**
64. Cheng F, Desai RJ, Handy DE, et al. 2018. Network-based approach to prediction and population-based validation of in silico drug repurposing. *Nat. Commun.* 9:2691. **https://doi.org/10.1038/s41467-018-05116-5**
65. Cheng F, Lu W, Liu C, et al. 2019. A genome-wide positioning systems network algorithm for in silico drug repurposing. *Nat. Commun.* 10:3476. **https://doi.org/10.1038/s41467-019-10744-6**
66. Zeng X, Zhu S, Lu W, et al. 2020. Target identification among known drugs by deep learning from heterogeneous networks. *Chem. Sci.* 11:1775–97. **https://doi.org/10.1039/C9SC04336E**
67. Cummings J, Zhou Y, Lee G, et al. 2024. Alzheimer's disease drug development pipeline: 2024. *Alzheimers Dement. Transl. Res. Clin. Interv.* 10:e12465. **https://doi.org/10.1002/trc2.12465**
68. Fang J, Zhang P, Wang Q, et al. 2022. Artificial intelligence framework identifies candidate targets for drug repurposing in Alzheimer's disease. *Alzheimers Res. Ther.* 14:7. **https://doi.org/10.1186/s13195-021-00951-z**
69. Hernan MA, Robins JM. 2016. Using big data to emulate a target trial when a randomized trial is not available. *Am. J. Epidemiol.* 183:758–64. **https://doi.org/10.1093/aje/kwv254**
70. Sanchez P, Voisey JP, Xia T, et al. 2022. Causal machine learning for healthcare and precision medicine. *R. Soc. Open Sci.* 9:220638. **https://doi.org/10.1098/rsos.220638**
71. Laifenfeld D, Yanover C, Ozery-Flato M, et al. 2021. Emulated clinical trials from longitudinal real-world data efficiently identify candidates for neurological disease modification: examples from Parkinson's disease. *Front. Pharmacol.* 12:631584. **https://doi.org/10.3389/fphar.2021.631584**
72. Lee S, Liu R, Cheng F, Zhang P. 2025. A deep subgrouping framework for precision drug repurposing via emulating clinical trials on real-world patient data. In *Proceedings of the 31st ACM SIGKDD Conference*

on Knowledge Discovery and Data Mining, Vol. 1. Association for Computing Machinery. **https://doi.org/10.1145/3690624.3709418**
73. Zang C, Zhang H, Xu J, et al. 2023. High-throughput target trial emulation for Alzheimer's disease drug repurposing with real-world data. *Nat. Commun.* 14:8180. **https://doi.org/10.1038/s41467-023-43929-1**
74. Yang Z, Huang Y, Jiang Y, et al. 2018. Clinical assistant diagnosis for electronic medical record based on convolutional neural network. *Sci. Rep.* 8:6329. **https://doi.org/10.1038/s41598-018-24389-w**
75. Rasmy L, Nigo M, Kannadath BS, et al. 2022. Recurrent neural network models (CovRNN) for predicting outcomes of patients with COVID-19 on admission to hospital: model development and validation using electronic health record data. *Lancet Digit. Health* 4:e415–25. **https://doi.org/10.1016/S2589-7500(22)00049-8**
76. Choi E, Xu Z, Li Y, et al. 2020. Learning the graphical structure of electronic health records with graph convolutional transformer. *Proc. AAAI Conf. Artificial Intell.* 34:606–13. **https://doi.org/10.1609/aaai.v34i01.5400**
77. Schildcrout JS, Heagerty PJ. 2008. On outcome-dependent sampling designs for longitudinal binary response data with time-varying covariates. *Biostatistics* 9:735–49. **https://doi.org/10.1093/biostatistics/kxn006**
78. Shao J, Zhong B. 2003. Last observation carry-forward and last observation analysis. *Stat. Med.* 22:2429–41. **https://doi.org/10.1002/sim.1519**
79. Goodfellow IJ, Pouget-Abadie J, Mirza M, et al. 2014. Generative adversarial nets. In *Advances in Neural Information Processing Systems 27*, ed. Z Ghahramani, M Welling, C Cortes, et al. Curran Associates
80. Zheng D, Song X, Ma C, et al. 2020. DGL-KE: Training knowledge graph embeddings at scale. In *Proceedings of the 43rd International ACM SIGIR Conference on Research and Development in Information Retrieval*. Association for Computing Machinery. **https://doi.org/10.1145/3397271.3401172**
81. Huang K, Chandak P, Wang Q, et al. 2024. A foundation model for clinician-centered drug repurposing. *Nat. Med.* 30:3601–13. **https://doi.org/10.1038/s41591-024-03233-x**
82. Zeng X, Song X, Ma T, et al. 2020. Repurpose open data to discover therapeutics for COVID-19. *J. Proteome Res.* 19:4624–36. **https://doi.org/10.1021/acs.jproteome.0c00316**
83. Xu J, Song W, Xu Z, et al. 2025. Single-microglia transcriptomic transition network-based prediction and real-world patient data validation identifies ketorolac as a repurposable drug for Alzheimer's disease. *Alzheimers Dement.* 21:e14373. **https://doi.org/10.1002/alz.14373**
84. Venugopalan J, Tong L, Hassanzadeh HR, Wang MD. 2021. Multimodal deep learning models for early detection of Alzheimer's disease stage. *Sci. Rep.* 11:3254. **https://doi.org/10.1038/s41598-020-74399-w**
85. Koutsouleris N, Dwyer DB, Degenhardt F, et al. 2021. Multimodal machine learning workflows for prediction of psychosis in patients with clinical high-risk syndromes and recent-onset depression. *JAMA Psychiatry* 78:195–209. **https://doi.org/10.1001/jamapsychiatry.2020.3604**
86. Zhavoronkov A, Ivanenkov YA, Aliper A, et al. 2019. Deep learning enables rapid identification of potent DDR1 kinase inhibitors. *Nat. Biotechnol.* 37:1038–40. **https://doi.org/10.1038/s41587-019-0224-x**
87. Sun M, Zhao S, Gilvary C, et al. 2020. Graph convolutional networks for computational drug development and discovery. *Brief Bioinform.* 21:919–35. **https://doi.org/10.1093/bib/bbz042**
88. Born J, Manica M. 2021. Trends in deep learning for property-driven drug design. *Curr. Med. Chem.* **https://doi.org/10.2174/0929867328666210729115728**
89. Cummings J, Zhou Y, Lee G, et al. 2025. Alzheimer's disease drug development pipeline: 2025. *Alzheimers Dement. Transl. Res. Clin. Interv.* 11:e70098. **https://doi.org/10.1002/trc2.70098**
90. Rodriguez S, Hug C, Todorov P, et al. 2021. Machine learning identifies candidates for drug repurposing in Alzheimer's disease. *Nat. Commun.* 12:1033. **https://doi.org/10.1038/s41467-021-21330-0**
91. Zhou Y, Fang J, Bekris LM, et al. 2021. AlzGPS: a genome-wide positioning systems platform to catalyze multi-omics for Alzheimer's drug discovery. *Alzheimers Res. Ther.* 13:24. **https://doi.org/10.1186/s13195-020-00760-w**
92. Xu J, Mao C, Hou Y, et al. 2022. Interpretable deep learning translation of GWAS and multi-omics findings to identify pathobiology and drug repurposing in Alzheimer's disease. *Cell Rep.* 41:111717. **https://doi.org/10.1016/j.celrep.2022.111717**

93. Xu J, Zhang P, Huang Y, et al. 2021. Multimodal single-cell/nucleus RNA sequencing data analysis uncovers molecular networks between disease-associated microglia and astrocytes with implications for drug repurposing in Alzheimer's disease. *Genome Res.* 31:1900–12. **https://doi.org/10.1101/gr.272484.120**
94. Shin MK, Vazquez-Rosa E, Koh Y, et al. 2021. Reducing acetylated tau is neuroprotective in brain injury. *Cell* 184:2715–32.e23. **https://doi.org/10.1016/j.cell.2021.03.032**
95. Min SW, Chen X, Tracy TE, et al. 2015. Critical role of acetylation in tau-mediated neurodegeneration and cognitive deficits. *Nat. Med.* 21:1154–62. **https://doi.org/10.1038/nm.3951**
96. Taubes A, Nova P, Zalocusky AK, et al. 2021. Experimental and real-world evidence supporting the computational repurposing of bumetanide for APOE4-related Alzheimer's disease. *Nat. Aging* 18:932–47. **https://doi.org/10.1038/s43587-021-00122-7**
97. Yahya EB, Alqadhi AM. 2021. Recent trends in cancer therapy: a review on the current state of gene delivery. *Life Sci.* 269:119087. **https://doi.org/10.1016/j.lfs.2021.119087**
98. Siegel RL, Giaquinto AN, Jemal A. 2024. Cancer statistics, 2024. *CA Cancer J. Clinic.* 74:12–49. **https://doi.org/10.3322/caac.21820**. Erratum. 2024. *CA Cancer J. Clinic.* 74:203
99. Tanoli Z, Vähä-Koskela M, Aittokallio T. 2021. Artificial intelligence, machine learning, and drug repurposing in cancer. *Expert Opin. Drug Discov.* 16:977–89. **https://doi.org/10.1080/17460441.2021.1883585**
100. Xu H, Aldrich MC, Chen Q, et al. 2015. Validating drug repurposing signals using electronic health records: a case study of metformin associated with reduced cancer mortality. *J. Am. Med. Inform. Assoc.* 22:179–91. **https://doi.org/10.1136/amiajnl-2014-002649**
101. Velavan TP, Meyer CG. 2020. The COVID-19 epidemic. *Trop. Med. Int. Health* 25:278–80. **https://doi.org/10.1111/tmi.13383**
102. Zhou Y, Wang F, Tang J, et al. 2020. Artificial intelligence in COVID-19 drug repurposing. *Lancet Digit. Health* 2:e667–76. **https://doi.org/10.1016/S2589-7500(20)30192-8**
103. Zhou Y, Hou Y, Shen J, et al. 2020. Network-based drug repurposing for novel coronavirus 2019-nCoV/SARS-CoV-2. *Cell Discov.* 6:14. **https://doi.org/10.1038/s41421-020-0153-3**
104. Zhou Y, Hou Y, Shen J, et al. 2020. A network medicine approach to investigation and population-based validation of disease manifestations and drug repurposing for COVID-19. *PLOS Biol.* 18:e3000970. **https://doi.org/10.1371/journal.pbio.3000970**
105. Martin WR, Cheng F. 2020. Repurposing of FDA-approved toremifene to treat COVID-19 by blocking the spike glycoprotein and NSP14 of SARS-CoV-2. *J. Proteome Res.* 19:4670–77. **https://doi.org/10.1021/acs.jproteome.0c00397**
106. Nabel EG. 2003. Cardiovascular disease. *New Engl. J. Med.* 349:60–72. **https://doi.org/10.1056/NEJMra035098**. Correction. 2003. *New Engl. J. Med.* 349:629
107. Paci P, Fiscon G, Conte F, et al. 2022. Comprehensive network medicine-based drug repositioning via integration of therapeutic efficacy and side effects. *npj Syst. Biol. Appl.* 8:12. **https://doi.org/10.1038/s41540-022-00221-0**
108. Jiménez ÁB, Lázaro JL, Dorronsoro JR. 2007. Finding optimal model parameters by discrete grid search. In *Innovations in Hybrid Intelligent Systems*, ed. E Corchado, JM Corchado, A Abraham. Springer. **https://doi.org/10.1007/978-3-540-74972-1_17**
109. He X, Zhao K, Chu X. 2021. AutoML: a survey of the state-of-the-art. *Knowl.-Based Syst.* 212:106622. **https://doi.org/10.1016/j.knosys.2020.106622**
110. Sandmann S, Riepenhausen S, Plagwitz L, Varghese J. 2024. Systematic analysis of ChatGPT, Google search and Llama 2 for clinical decision support tasks. *Nat. Commun.* 15:2050. **https://doi.org/10.1038/s41467-024-46411-8**
111. Zhang Y, Li Y, Cui L, et al. 2023. Siren's song in the AI ocean: a survey on hallucination in large language models. Preprint, arXiv:2309.01219v2 [cs.CL]
112. Swanson K, Wu W, Bulaong NL, et al. 2024. The virtual lab: AI agents design new SARS-CoV-2 nanobodies with experimental validation. Preprint, bioRxiv. **https://www.biorxiv.org/content/10.1101/2024.11.11.623004v1**
113. Su X, Wang Y, Gao S, et al. 2024. Knowledge graph based agent for complex, knowledge-intensive QA in medicine. Preprint, arXiv:2410.04660v1

114. Abramson J, Adler J, Dunger J, et al. 2024. Accurate structure prediction of biomolecular interactions with AlphaFold 3. *Nature* 630:493–500. **https://doi.org/10.1038/s41586-024-07487-w**
115. Nebel RA, Aggarwal NT, Barnes LL, et al. 2018. Understanding the impact of sex and gender in Alzheimer's disease: a call to action. *Alzheimers Dement.* 14:1171–83. **https://doi.org/10.1016/j.jalz.2018.04.008**
116. Mehta KM, Yaffe K, Perez-Stable EJ, et al. 2008. Race/ethnic differences in AD survival in US Alzheimer's Disease Centers. *Neurology* 70:1163–70. **https://doi.org/10.1212/01.wnl.0000285287.99923.3c**
117. Raman R, Quiroz YT, Langford O, et al. 2021. Disparities by race and ethnicity among adults recruited for a preclinical Alzheimer disease trial. *JAMA Netw. Open* 4:e2114364. **https://doi.org/10.1001/jamanetworkopen.2021.14364**
118. Braunstein ML. 2018. Healthcare in the age of interoperability: the promise of fast healthcare interoperability resources. *IEEE Pulse* 9:24–27. **https://doi.org/10.1109/MPUL.2018.2869317**
119. Xu J, Glicksberg BS, Su C, et al. 2020. Federated learning for healthcare informatics. *J. Healthc. Inform. Res.* 5:1–19. **https://doi.org/10.1007/s41666-020-00082-4**
120. Wang F, Kaushal R, Khullar D. 2020. Should health care demand interpretable artificial intelligence or accept "black box" medicine? *Ann. Intern. Med.* 172:59–60. **https://doi.org/10.7326/M19-2548**
121. Martin W, Sheynkman G, Lightstone FC, et al. 2021. Interpretable artificial intelligence and exascale molecular dynamics simulations to reveal kinetics: applications to Alzheimer's disease. *Curr. Opin. Struct. Biol.* 72:103–13. **https://doi.org/10.1016/j.sbi.2021.09.001**
122. Alkhulaifi A, Alsahli F, Ahmad I. 2021. Knowledge distillation in deep learning and its applications. *PeerJ Comput. Sci.* 7:e474. **https://doi.org/10.7717/peerj-cs.474**
123. Piñero J, Bravo À, Queralt-Rosinach N, et al. 2017. DisGeNET: a comprehensive platform integrating information on human disease-associated genes and variants. *Nucleic Acids Res.* 45:D833–39. **https://doi.org/10.1093/nar/gkw943**
124. Stelzer G, Rosen N, Plaschkes I, et al. 2016. The GeneCards suite: from gene data mining to disease genome sequence analyses. *Curr. Protoc. Bioinform.* 54:1.30.31–33. **https://doi.org/10.1002/cpbi.5**
125. Ochoa D, Hercules A, Carmona M, et al. 2021. Open Targets Platform: supporting systematic drug–target identification and prioritisation. *Nucleic Acids Res.* 49:D1302–10. **https://doi.org/10.1093/nar/gkaa1027**
126. Howe KL, Achuthan P, Allen J, et al. 2021. Ensembl 2021. *Nucleic Acids Res.* 49:D884–91. **https://doi.org/10.1093/nar/gkaa942**
127. Crystallography: Protein Data Bank. 1971. *Nat. New Biol.* 233:223
128. Meyer MJ, Beltran JF, Liang S, et al. 2018. Interactome INSIDER: a structural interactome browser for genomic studies. *Nat. Methods* 15:107–14. **https://doi.org/10.1038/nmeth.4540**
129. Fabregat A, Jupe S, Matthews L, et al. 2018. The Reactome Pathway Knowledgebase. *Nucleic Acids Res.* 46:D649–55. **https://doi.org/10.1093/nar/gkx1132**
130. Szklarczyk D, Gable AL, Lyon D, et al. 2019. STRING v11: protein-protein association networks with increased coverage, supporting functional discovery in genome-wide experimental datasets. *Nucleic Acids Res.* 47:D607–13. **https://doi.org/10.1093/nar/gky1131**
131. Zeng X, Zhu S, Liu X, et al. 2019. deepDR: a network-based deep learning approach to *in silico* drug repositioning. *Bioinformatics* 35:5191–98. **https://doi.org/10.1093/bioinformatics/btz418**
132. Yang Y, Qiu Y, Hu J, et al. 2024. A deep learning framework combining molecular image and protein structural representations identifies candidate drugs for pain. *Cell Rep. Methods* 4:100865. **https://doi.org/10.1016/j.crmeth.2024.100865**
133. Wang Y, Wang J, Cao Z, Barati Farimani A. 2022. Molecular contrastive learning of representations via graph neural networks. *Nat. Mach. Intell.* 4:279–87. **https://doi.org/10.1038/s42256-022-00447-x**