

Review

Navigating the Frontiers of Machine Learning in Neurodegenerative Disease Therapeutics

Yoonjeong Cha, Mohamedi N. Kagalwala and Jermaine Ross *

Alleo Labs, San Francisco, CA 94105, USA; yj@alleolabs.com (Y.C.); mo@alleolabs.com (M.N.K.)

* Correspondence: ross@alleolabs.com

Abstract: Recent advances in machine learning hold tremendous potential for enhancing the way we develop new medicines. Over the years, machine learning has been adopted in nearly all facets of drug discovery, including patient stratification, lead discovery, biomarker development, and clinical trial design. In this review, we will discuss the latest developments linking machine learning and CNS drug discovery. While machine learning has aided our understanding of chronic diseases like Alzheimer's disease and Parkinson's disease, only modest effective therapies currently exist. We highlight promising new efforts led by academia and emerging biotech companies to leverage machine learning for exploring new therapies. These approaches aim to not only accelerate drug development but to improve the detection and treatment of neurodegenerative diseases.

Keywords: machine learning; drug discovery; multiple sclerosis; amyotrophic lateral sclerosis; Parkinson's disease; Alzheimer's disease; Huntington's disease

1. Introduction

The average investment in bringing a new drug to market ranges between \$314 million and \$2.8 billion, spanning over 10 to 15 years [1,2]. Despite rapid innovations in biotechnology equipment aimed at reducing the operating costs, drug development is primarily dependent on classical methods for assessing the safety and efficacy of drug candidates [3,4]. Such methods are associated with a number of pain points, including but not limited to patient stratification, target identification, high-throughput screening, drug design and optimization, biomarker discovery, and clinical trial design. Each exercise often employs an expensive brute-force approach that would largely be overlooked if not for the high attrition rates in drug development: oncology drugs have a 3.4% overall probability of success of gaining approval from the United States Food and Drug Administration (FDA); neuroscience drugs 15%; autoimmune/inflammation drugs, 15.1%; cardiovascular drugs, 25.5%; and vaccines 33.4% [5]. In this review, we will explore recent progress in machine learning (ML) that has enabled innovative approaches along the pipeline of CNS drug discovery. These promising methodologies seek to improve not only the cost and timelines of drug development but also the efficacy of drugs targeting neurodegenerative diseases.

2. Currently Approved Treatments for Neurodegeneration

The most common neurodegenerative diseases include Alzheimer's disease (AD), Parkinson's disease (PD), multiple sclerosis (MS), amyotrophic lateral sclerosis (ALS), and Huntington's disease (HD) [6,7]. While significant progress has been made in understanding these disorders, there have been a limited number of effective treatments developed to modify the disease progression and severity in patients. For example, Alzheimer's disease is the seventh leading cause of death in the United States and accounts for the greatest number of dementia cases worldwide [8]. In the early 1980s, a US-based workgroup at the National Institute of Neurological and Communicative Disorders and Stroke (currently known as the National Institute of Neurological Disorders and Stroke) established an



Citation: Cha, Y.; Kagalwala, M.N.; Ross, J. Navigating the Frontiers of Machine Learning in Neurodegenerative Disease Therapeutics. *Pharmaceuticals* **2024**, *17*, 158. <https://doi.org/10.3390/ph17020158>

Academic Editors: Weida Tong and Rebecca Kusko

Received: 30 December 2023

Revised: 16 January 2024

Accepted: 23 January 2024

Published: 25 January 2024



Copyright: © 2024 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

universally accepted criteria for the diagnosis and staging of AD which triggered a modern age of research in the field [9]. Despite four decades of rigorous scientific effort, only seven treatments for AD have been approved by the FDA, with only two new treatments, Aducanumab and Lecanemab, since 2003. Five of the approved treatments, including galantamine (Razadyne), rivastigmine (Exelon), donepezil (Aricept), memantine (Namenda), and memantine/donepezil (Namzaric), are widely considered to only briefly and modestly improve AD symptoms, ultimately failing to prevent or slow disease progression [10]. Similarly, there are only seven approved drugs for ALS, including tofersen (Qalsody), sodium phenylbutyrate/taurursodiol (Relyvrio), edaravone (Radicava), riluzole (Rilutek, Tiglutik, Exservan), and dextromethorphan/quinidine (Nuedexta) [11]. Like the approved AD medications, most of these ALS therapies do not reverse or stop progression but instead relieve symptoms or delay progression in patients [12,13].

3. Link between Heterogeneity and Novel Disease Targets in Neurological Disorders

3.1. Genetic Heterogeneity

While the complexity of neurological disorders may partly explain the lack of success in drug development in this field, there is a growing amount of evidence supporting heterogeneity among patients with AD [14–18], ALS [19–21], and PD [22–24]. For sporadic forms of common neurological disorders, clinical diagnosis has been broadly applied, where patients present varying clinical features, including but not limited to disease onset and progression, symptomology, and clinical outcome. However, standardized criteria for neuropsychological assessment have proven often insufficient for differential diagnosis, and the lack of robust biomarkers has complicated diagnostic and prognostic work-up for neurological cases [25]. Genetic studies have provided clarity on the causative mutations in familial forms of Alzheimer's disease, where characterized variants in the amyloid precursor protein (APP), presenilin 1 (PSEN1), and presenilin 2 (PSEN2) genes have been shown to be nearly but not fully penetrant [26], whereas familial Parkinson's mutations in genes like leucine-rich repeat kinase 2 (LRRK2), glucocerebrosidase (GBA), Parkin (PRKN), and alpha-synuclein (SCNA) have been useful for determining PD risk, diagnosis, and disease progression [27]. Despite extensive knowledge of the genetic factors among risk carriers, the clinical heterogeneity among these cases is not completely understood [28].

3.2. Publicly Available Repositories for Deciphering the Heterogeneity within Neurodegeneration

An initial step toward deciphering the heterogeneity of neurodegenerative diseases may require stratifying patients into distinct cohorts based on biological data. Over the past decade, several comparative studies have expanded access to rich neurodegenerative datasets derived from medical imaging and biospecimen samples, including brain magnetic resonance imaging (MRI), positron emission tomography, postmortem brain and peripheral nerve tissue, cerebrospinal fluid, plasma, and electroencephalographs. Such publicly available repositories include the Alzheimer's Disease Neuroimaging Initiative (ADNI) [29], the Alzheimer's Disease Data Initiative (ADDI) [30], the Religious Orders Study and Rush Memory and Aging Project (ROS MAP) [31], the Accelerating Medicines Partnership Program for Alzheimer's Disease project (AMP-AD) [32], the Parkinson's Progression Markers Initiative (PPMI) [33], the Answer ALS project [34], and the Target ALS project [35], as well as others (Table 1). Arguably, these resources represent neurodegenerative-based counterparts to oncology-based data initiatives such as The Cancer Genome Atlas Program (commonly known as TCGA), which provides access to 2.5 petabytes of multi-omics data across 33 cancer types [36].

Table 1. Publicly available repositories with neurodegenerative disease data.

Data Repository	Disease Area	Data Types	Reference
Alzheimer’s Disease Neuroimaging Initiative	AD	Brain magnetic resonance imaging, positron emission tomography, multi-omics, clinical, fluid biomarkers	[29]
Alzheimer’s Disease Data Initiative	AD	Multi-omics, clinical trial readouts	[30]
Religious Orders Study and Rush Memory and Aging Project	AD	Multi-omics, brain magnetic resonance imaging, neuropathology, clinical, fluid biomarkers	[31]
Accelerating Medicines Partnership Program for Alzheimer’s Disease	AD, PD, other NDDs	Multi-omics, brain magnetic resonance imaging electrophysiology	[32]
Parkinson’s Progression Markers Initiative	PD	Multi-omics, brain magnetic resonance imaging, clinical	[33]
Answer ALS Project	ALS	Multi-omics, clinical	[34]
Target ALS Project	ALS	Multi-omics, clinical	[35]

Alzheimer’s, AD; amyotrophic lateral sclerosis, ALS; Parkinson’s, PD; neurodegeneration, NDD.

3.3. Computational Approaches to Stratifying Patients in Oncology

There are illustrative examples of employing transcriptomics and ML to subtype cancer patients based on biologically relevant associations, offering a starting point for applying similar approaches to classifying patients with neurodegenerative disorders. For example, a large body of literature has shed light on the genomic and epigenomic deregulation in cancer biology and its relationship to clinical heterogeneity. Breast cancer is widely known to be a highly heterogeneous disease, with differences observed across genomic, epigenomic, transcriptomic, and proteomic data [37]. Several bioinformatic approaches have been employed to unravel the patient stratification across different cancer patients like BRCA1/2 (breast cancer 1/2) mutation carriers. Notably, transcriptome analyses have revealed gene expression differences comparing BRCA1 and BRCA2 subjects as well as between breast and ovarian cancer patients [38]. Similarly, lung cancer studies have found considerable variation among histological samples associated with clinicopathological features [39], and gene expression analysis confirmed unique transcriptional profiles among lung adenocarcinoma and squamous cell carcinoma subtypes [40]. Further, classical ML methods, such as unsupervised learning, have demonstrated significant improvement in subclassifying tumors using gene expression data [40,41]. In particular, clinically relevant subtypes were characterized among luminal breast cancer samples by utilizing consensus clustering, an unsupervised ML technique that offers improvements in stability over the classical clustering methods [42–44].

3.4. Applications of ML to Stratifying Patients with Neurodegeneration

Beyond consensus clustering, more advanced ML algorithms have achieved comprehensive patient subtyping via the integration of diverse data types. Such examples encompass, but are not restricted to, Similarity Network Fusion (SNF), Pattern Fusion Analysis (PFA), NEMO (Neighborhood-Based Multi-Omics Clustering), non-negative matrix factorization (NMF), Subtype-GAN, and Perturbation Clustering for Data Integration and Disease Subtyping (PINS). While NEMO, SNF, and PINS are primarily based on similarity networks, PFA and NMF are grounded in the principles of dimensionality reduction. Recently, NMF was employed to stratify a large cohort of ALS patients based on samples generated from the Target ALS project [45]. Clustering analysis revealed three unique ALS subgroups, which were defined by transcriptional differences in biologically relevant

mechanisms, including oxidative stress, reactive gliosis, and RNA dysregulation. ALS subtype patients associated with RNA dysregulation were linked to elevated levels of TAR DNA-binding protein 43 (TDP-43), a regulator of RNA processing known for its pathogenic role in ALS. Consistent with this notion, ALS subtype patients with RNA dysregulation exhibited initial limb symptoms, with prior research associating limb onset with TDP-43 pathology [46]. With an increasing body of evidence suggesting the involvement of RNA dysregulation in ALS [21], patient stratification presents an effective approach to discovering novel targets best suited for precision therapy.

4. Computational Approaches to Lead Discovery

4.1. Overview of ML in Lead Discovery

Lead discovery is an important stage in the drug discovery process. During this phase, chemical compounds, aimed against a specific target of interest, are identified and optimized to exert an ideal biological effect [47]. The latest research supports the potential of ML to improve the efficiency of pharmacological development. Concretely, drug hunters have applied ML to various bottlenecks of lead discovery, including hit-to-lead and lead optimization, and have developed approaches to the computational prediction of protein structures, virtual screening via structure-based/ligand-based methods, and the physicochemical optimization of lead drug candidates [47,48]. While generally agnostic to the disease area, the ML tools and methods described below have shown immense value in the field of CNS drug discovery.

4.2. Binding Site and Protein Structure Prediction

Computer-aided lead discovery starts with employing the available structural information on a disease target. Proteins are commonly studied as three-dimensional (or tertiary) structures, traditionally obtained using various prevalent methods such as X-ray crystallography, NMR spectroscopy, and cryo-electron microscopy [49]. The structural information is then pre-processed and analyzed to identify potential ligand-binding sites [50]. There are a number of existing algorithms available for binding pocket prediction (Table 2), including but not limited to Schrödinger's SiteMap [51], Fpocket [52], DoGSiteScorer [53], and Q-SiteFinder [54]. The foundation of these tools varies widely, encompassing diverse technologies aimed at achieving accurate prediction. For example, SiteMap employs a grid-based method to evaluate the free energy profiles and geometry of the putative ligand motifs present on a protein target, whereas Fpocket is largely restricted to resolving binding cavities based on geometry alone. While these techniques are frequently used throughout CNS drug discovery [55–57], Fpocket and SiteMap utilize computational geometry and physics-based principles as opposed to ML. In contrast, recent advances have applied convolutional neural networks (CNNs) to resolving the putative functional pockets within neurodegenerative proteins. A CNN is a neural network that detects patterns in the input data, such as amino acids in proteins or atomic symbols in compounds. DeepSite is a deep CNN trained on over 7000 protein structures curated from a publicly available annotated database called sc-PDB [58], which comprises binding sites characterized from protein structures found in the Protein Data Bank [59]. Recently, DeepSite analysis revealed allosteric binding motifs in a neuronal protein known as Synapsin III (Syn III) and highlighted the structural interaction between Syn III and methylphenidate, a monoamine reuptake inhibitor used for treating attention deficit hyperactivity disorder [60]. Syn III is a member of the synapsin protein family, a group of evolutionarily conserved phospho-proteins crucial for regulating synaptic transmitter release and facilitating neuronal communication, and has been recently associated with the aggregated α -synuclein found in PD and dementia with Lewy bodies (DLB) [61]. While relatively new compared to traditional prediction algorithms, deep CNNs have the potential to improve or complement geometry-based and physics-based predictions of the ligand-binding sites characterized in experimentally validated protein structures.

Table 2. Selected computational tools available for predicting structure–activity relationships during CNS drug development.

Drug Discovery Application	Algorithm Examples	CNS Target Examples
Protein binding site prediction	SiteMap, Fpocket, DoGSiteScorer, Q-SiteFinder, DeepSite	Synapsin III
Protein structure prediction	RoseTTAFold, I-TASSER, AlphaFold, QUARK	PINK1, PSEN1, APP, APOE, TREM2
Ligand-based virtual screening	SwissSimilarity	α -synuclein
Structure-based virtual screening	DeepDTA, GraphDTA, DeepGS, 3-Tunnel DNN, AtomNet	Mfn2, GluA2

Due to the practical challenges associated with crystallography and NMR spectroscopy [62], there has been a growing trend in the accurate de novo prediction of protein structures using bioinformatics and ML (Table 2), including RoseTTAFold [63], I-TASSER [64], AlphaFold [65], and QUARK [66]. Similar to SiteMap, conventional predictions of protein structures rely on the principles governing protein energy functions—mathematical models that measure the energy linked to the shape or arrangement of a protein given a particular amino acid sequence. While energy-based modeling is computationally expensive, CPUs (Central Processing Units) and GPUs (Graphics Processing Units) have greatly improved over the years, which has yielded better predictions. In addition, parallelization and distributed computing have significantly increased the capacity to run simulations and computations at a large scale [67]. For example, NVIDIA’s CUDA (Compute Unified Device Architecture) has allowed developers to maximize the potential from GPUs for deep learning, including developing more effective force fields used in energy-based modeling [68].

AlphaFold and RoseTTAFold have become two widely adopted tools for modern protein structure prediction. Both deep learning tools can be used for ab initio folding, which is a method for predicting protein structures based solely on amino acid sequences. Conversely, template-based methods leverage existing experimental structure data to make de novo predictions. While RoseTTAFold combines both template-based modeling and ab initio folding, AlphaFold largely depends on ab initio folding, with lesser importance of the templates. Both AlphaFold and RoseTTAFold have been used to study the protein targets associated with neurodegeneration. PINK1 (PTEN-induced putative kinase 1) is a serine/threonine kinase known for its role in mitophagy and its impact on AD, ALS, HD, and PD [69]. Structure and mutagenesis studies have revealed disease-linked mutations within the functional kinase domain of PINK1, including at the 288th amino acid position—a serine residue (Ser288) crucial to autophosphorylation [70]. In contrast, there are several PINK1 mutations located in regions not included in the published structures of PINK1 [70]. AlphaFold analysis revealed the complete structure of human PINK1, including the presence of an alpha helix in the N-terminal region. Confirmed using mass spectrometry, the domain was subsequently shown to be necessary for Ser228 autophosphorylation and PINK1 activation, exhibiting a potential therapeutic mechanism in PINK1 patients [70]. Comparisons between AlphaFold and RoseTTAFold have also been conducted in structural prediction studies. Genome-wide association studies have revealed disease-associated mutations in PSEN1, APP, APOE (Apolipoprotein E), and TREM2 (Triggering Receptor Expressed on Myeloid Cells 2) [71], well-studied proteins that are the focus of therapeutic intervention for AD [72–75]. Protein structure predictions were carried out for all four proteins using both AlphaFold and RoseTTAFold to assess the accuracy of each modeling method against experimentally validated structures. The benchmark performance was evaluated using two metrics for structural similarity: Root Mean Square Deviation (RMSD) and Template Modeling score (TM-score). RMSD is frequently more effective in capturing the overall general structural similarity, even in instances where no experimental struc-

ture reference is available (or ab initio predictions) [76]. The TM-score is arguably more robust than RMSD as it considers the entire structural alignment, enabling it to detect finer structural variations, although this metric is more applicable to template-based predictions [76]. The TM-scores and RMSD estimates revealed a high degree of similarity between AlphaFold and RoseTTAFold when predicting the protein structures of PSEN1, APP, APOE, and TREM2 [77]. The modeling predictions also resolved gaps in PSEN1 that were not captured using X-ray crystallography or cryo-electron microscopy [77]. Generally, X-ray crystallography is not very sensitive to the mobility of proteins, whereas template-based predictions may be a powerful tool for elucidating the intrinsically flexible segments of partially captured proteins.

4.3. Hit Identification via Virtual Screening

During the hit identification stage of drug discovery, a proven method for identifying chemical hits to targets includes purifying the disease target proteins, establishing robust biochemical assays, conducting high-throughput screening (HTS) of chemical libraries, and separating out active hits [78]. Although automation and miniaturization have contributed to reducing costs, the well-established practice remains relatively expensive and inefficient, particularly for novel and higher-risk targets, as the costs are tied to the size of the chemical library and its scope [79]. To mitigate risks, there has been a greater focus on utilizing ML for binding prediction between targets and ligands. In broad terms, there are two approaches to virtual HTS: ligand-based screening and structure-based screening. Structure-based screening can be further divided into complex-based and pair-based models [80]. While ligand-based screening typically relies on similarity measures, the majority of structure-based models integrate some application of ML. For structure-based screening, quantitative structure–activity relationships (SARs) can be predicted with or without employing a tertiary structure, which, until recently, was largely confined to experimentally resolved structures, but now, with advancements like the AlphaFold and RoseTTAFold algorithms, has seen expanded possibilities. Instead of tertiary structures, pair-based screening involves training models using primary representations of proteins in the form of SMILES as input, coupled with biochemical activity data, which benefit from being more deployable compared to complex-based screening [81]. Specifically, training billions of compounds using deep CNNs based on SMILES representation is significantly less computationally expensive relative to physics-based 3D docking methods.

Recently, virtual ligand-based HTS was utilized to discover inhibitors targeting the α -synuclein protein [56]. α -synuclein is a pathological hallmark of PD and DLB, and its aggregation is associated with the degeneration of the dopaminergic neurons residing in the substantia nigra pars compacta, a brain region involved in motor planning [56]. To identify α -synuclein binders, a virtual HTS was performed using SwissSimilarity (Table 2), a web-accessible tool for identifying putative hits using a diverse collection of promising and validated small molecule libraries including but not limited to 3071 approved drugs and 2989 drug candidates from the ChEMBL 29 database, over 320,000 commercially available molecules from the SPECS library, over 9 million molecules from the ZINC20 database, and over 30 million molecules from the Enamine “REAL” catalog [82]. To prime the search, SwissSimilarity analysis was conducted using two known α -synuclein binders, namely SynuClean-D and ZPD-2. Both α -synuclein hits represent tool compounds validated in cell-based experiments but lack drug-like properties [83]. Ligand-based screening using the SPECS library revealed analogs of SynuClean-D and ZPD-2, which were selected by leveraging multiple molecular fingerprinting methods. Each of these approaches facilitate similarity analysis by offering a distinct representation of the molecular structures, such as electrostatic properties, predefined chemical substructures, and the distribution of atomic charge [84]. The SwissSimilarity analysis identified 363 putative analogs of SynuClean-D and ZPD-2, which were further filtered for ideal drug-like properties, the absence of PAINS (pan-assay interference compounds), and commercial availability. A final set of 34 structures was selected for experimental validation, including a structurally similar

analog denoted as MeSC-04. Cell-based assays showed that MeSC-04 is a potent inhibitor of α -synuclein amyloid formation. Fpocket and SiteMap were employed to identify the binding pockets of α -synuclein, and molecular docking was performed to evaluate the binding interaction between MeSC-04 and the identified motifs. The molecular docking studies demonstrated binding interactions consistent with the previously reported findings involving SynuClean-D and α -synuclein, supporting the utility of ligand-based screening for chemical hits [56].

Pair-based screening is focused on predicting the quantitative SAR in protein–protein or protein–ligand interactions independent of knowing the native structure of the proteins or the ligands [80]. Ligands are inputted as SMILES, molecular fingerprints, or molecular graphs, whereas proteins are represented using full or partial sequences of amino acids. Pair-based screening can be performed using random forests, support vector machines, multilayer perceptrons, and neural networks [80]. Advanced architectures commonly utilize different types of neural networks, specifically recurrent neural networks, deep CNNs, and graph CNNs. Several CNN-based applications, such as DeepDTA and GraphDTA, are open-source and available for performing pair-based screening of ligand libraries (Table 2). For example, DeepDTA was recently utilized to identify hits for Mitofusin-2 (Mfn2), a GTPase associated with mitochondrial dysfunction that is implicated in the underlying pathology of AD [85]. Mfn2 is one of two paralogs of the mitofusin protein family, which are primarily responsible for the fusion of mitochondrial outer membranes [86]. DeepDTA was trained on a protein–ligand binding affinity database that consists of 1063 approved drugs called the Davis dataset, and its performance was compared to other models (Table 2), including GraphDTA (a graph CNN), DeepGS (a deep CNN), and a novel architecture called a three-tunnel deep neural network (a deep CNN denoted as 3-Tunnel DNN). To improve the training on the protein–drug binding affinity, a 3-Tunnel DNN distinguishes itself from other deep CNN models by explicitly integrating information from both positive samples (indicating protein–drug interactions) and negative samples (representing the absence of interactions), as well as incorporating protein sequences. When assessing the training performance, all tested models demonstrated comparable benchmarks, evaluated using metrics such as mean square error and the consistency index. Analysis of the 3-Tunnel DNN model screening revealed several approved drugs that exhibit potential to be repurposed for the inhibition of Mfn2 activity. Notably, Lamotrigine, Bosentan, Fluphenazine, Nabumetone, and Carbamazepine, featured in the leading drug hit list, are all medications previously investigated for their potential in AD treatment [85].

Complex-based screening aims to predict quantitative SARs in protein–protein or protein–ligand interactions by utilizing structural information on both the proteins and ligands [80]. Similar to pair-based models, complex-based methods consist of classical and modern ML approaches, commonly incorporating complex neural networks and encoding proteins as 3D grids [80,87]. For example, recent applications have employed deep CNNs for the structure-based screening of ligands against neurodegenerative protein targets, such as AMPA (α -amino-3-hydroxy-5-methyl-4-isoxazolepropionic acid) receptors. AMPA receptors are widely expressed in the CNS, and their dysfunction likely mediates the glutamate excitotoxicity underlying neuronal death and disease progression in MS [88]. A recent study utilized a deep-CNN-guided approach to identifying hits that may bind to an allosteric pocket located on one of the four subunits of AMPA receptors, known as glutamate receptor 2 (GluA2) [89], notable for its role in the regulation of Ca^{2+} permeation and voltage rectification [90]. Complex-based screening was conducted utilizing Atomwise's proprietary CNN (AtomNet) for predict the binding affinity of small molecules to GluA2 (Table 2) [90]. The effects of 50 putative GluA2 hits were validated using a cell-based assay to assess the glutamate-mediated excitotoxicity. In vitro models confirmed that glutamate-mediated excitotoxicity was inhibited by several hits, including highly potent compounds denoted as YH668, ZCAN155, and ZCAN262 [90]. Pharmacokinetic studies revealed that ZCAN262 also had good oral bioavailability and brain exposure. Animal

studies demonstrated that ZCAN262 treatment is sufficient to rescue myelination and axon integrity in EAE mice, an MS mouse model [90].

4.4. Lead Optimization Using ML

After hit identification, drug discovery teams have often embarked on intensive campaigns of medicinal chemistry to characterize drug candidates for Investigational New Drug (IND)-enabling studies. Such efforts of drug discovery can be broken down into the following stages: hit-to-lead, lead identification, and lead optimization. All phases of development involve rapid analog generation to improve their physicochemical properties and advance potential leads toward having drug-like characteristics. Collectively, the objective is to satisfy a set of predefined requirements known as a Target Product Profile (TPP) [91,92]. While context-dependent, the TPP broadly consists of thresholds for safety and efficacy. Concretely, the focus is on optimizing the parameters for ADME (absorption, distribution, metabolism, and excretion) to improve the overall bioavailability and target engagement while also attempting to reduce any safety/toxicity liabilities. Examples of ADME properties include aqueous solubility, membrane permeability, microsomal stability, and blood–brain barrier (BBB) penetrance, whereas early safety and toxicity assessment evaluates the inhibition of hERG (the human Ether-à-go-go gene) and CYP (cytochrome P450) activity [92].

To accelerate the discovery of CNS drug candidates, ML approaches have been developed to predict the optimization of ADME and toxicity. For example, DeePred-BBB is a deep CNN for predicting BBB permeability [93]. DeePred-BBB was trained on a broad set of 3605 compounds screened for BBB permeability and was benchmarked using the area under the curve (AUC). Compared to other published BBB permeability prediction models, DeePred-BBB performed relatively well with an AUC of 0.992. In contrast, the best reported AUC from another model is 0.98, which also employed a deep learning approach but was trained on a relatively smaller BBB dataset (462 compounds) [93]. Beyond DeePred-BBB, there are other emerging ML solutions for lead optimization, including those that incorporate generative artificial intelligence (AI) tools like large language models (LLMs), i.e., Bidirectional Encoder Representations from Transformers (BERT). For example, Mol-BERT was trained on datasets to predict not only BBB permeability but also clinical toxicity [94]. Applications like DeePred-BBB and Mol-BERT represent a promising new era of ML-guided drug design.

5. Industry Case Studies

Over the years, several biotech companies have emerged with a focus on using cutting-edge ML approaches for CNS drug discovery (Figure 1). From target identification to clinical trial design, these biotech companies have leveraged ML to accelerate therapeutic discovery, rapidly establishing drug pipeline programs and state-of-the-art platform technologies. For example, several companies, such as Verge Genomics [95], Alleo Labs [96], Insitro [97], Evotec [98], InveniAI [99], and Recursion [100], have pioneered the development of ML platforms for CNS target identification. Meanwhile, Schrödinger, an industry leader in complex-based screening, recently partnered with Otsuka Pharmaceutical and Bristol Myers Squibb to perform hit identification and lead optimization for potential CNS therapies [101]. Similar to WaveBreak Therapeutics [102] and BenevolentAI [103,104], Vincere Biosciences is applying GPU-powered ML to lead discovery using its own proprietary software for screening and optimizing small molecules [105]. Currently, Vincere is actively pursuing inhibitors for USP30, a deubiquitinating (DUB) enzyme implicated in PD. Alleo Labs, a biotech developing ML-guided precision medicine, is employing LLMs for small-molecule optimization of lead inhibitors for novel AD and PD targets, including DUB enzymes [106]. AbbVie and BigHat Biosciences recently formed a collaboration to leverage BigHat's ML design platform for treatments in neuroscience [107]. BigHat's platform employs the principles underlying generative AI to characterize and optimize antibodies [108]. Verge Genomics has utilized its end-to-end ML technology to identify novel ALS targets

and develop small-molecule therapeutics, namely VRG50635, an inhibitor of kinase PIKfyve (also known as Phosphoinositide Kinase, FYVE-Type Zinc-Finger-Containing). Verge successfully evaluated VRG50635 for its safety and tolerability in phase 1 clinical trials [109]. Beyond therapeutics, several biotech companies, such as NeuBio [110], Perceiv AI [111], Rune Labs [112], and LinusBio [113], are focusing on identifying robust biomarkers, as well as optimizing clinical trial design. Concretely, NeuBio is seeking to develop a blood test that can accurately diagnose disease in the earliest stages of development of neurodegeneration, by analyzing publicly available transcriptomic datasets from case–control studies of prodromal and early-stage disease using an evolutionary ML platform. NeuBio has assembled a panel of 141 RNA-based biomarkers that can be used for accurate diagnosis of AD, PD, and ALS. To inform patient selection and stratification when designing clinical trials for AD, Perceiv AI has developed a predictive ML platform that integrates different data types, such as fluid, genetic, and imaging biomarkers. In addition to advancing the field of artificial intelligence, NVIDIA has played a pivotal role in supporting ML-based biotech startups, such as Alleo Labs, Vincere, and Perceiv AI, through the NVIDIA Inception Program [114].

Companies Focused on ML in CNS Drug Discovery

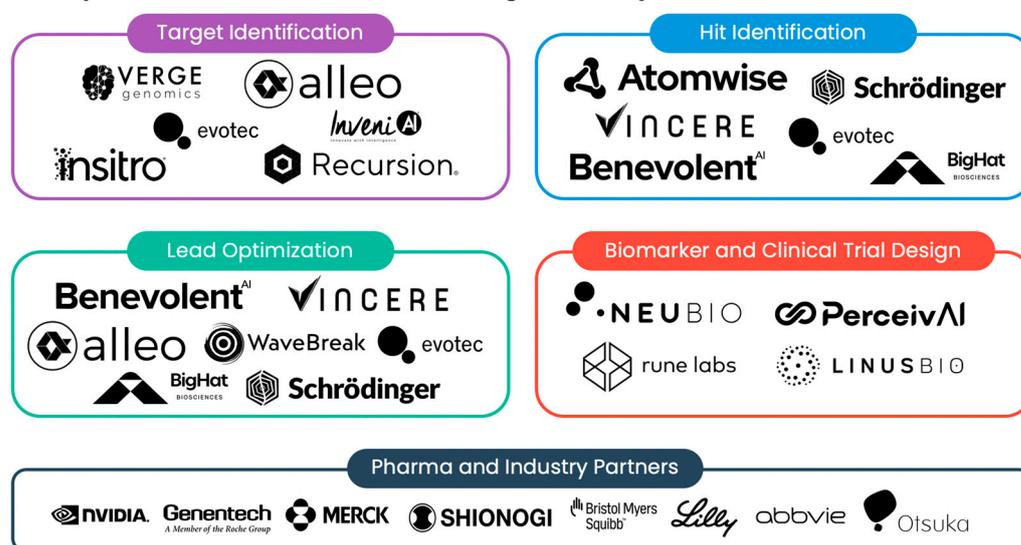


Figure 1. Selected companies using machine learning in CNS drug discovery. Illustrated is a summary of the biotech companies leveraging machine learning (ML) across different domains, including target identification, hit identification, lead optimization, biomarker design, and clinical trial design. Partner pharmaceutical and industry companies are also shown.

6. Conclusions

In this review, we summarized the advanced ML tools and approaches employed at various stages of CNS drug discovery. Given that patient stratification may be required to investigate new targets and treatments for neurodegeneration, we noted the utility of leveraging modern clustering algorithms to subtype patients using biological data, including the resources available via existing online repositories. We also examined several examples of employing more sophisticated neural networks to identify and design treatments during lead discovery. Lastly, we illustrated ongoing efforts to utilize ML for improving the clinical study design in neurodegenerative diseases. As these tools evolve, ML shows significant potential in reshaping the field of CNS drug discovery.

Author Contributions: This work was a collaborative effort amongst all authors. Y.C., M.N.K., and J.R. wrote parts of the manuscript. All authors assisted in editing the paper and contributed to the structure of this project and provided critical references. All authors have read and agreed to the published version of the manuscript.

Funding: The authors acknowledge financial support from the Alzheimer’s Disease Data Initiative and the NVIDIA Inception Program.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Not applicable.

Conflicts of Interest: Jermaine Ross is an employee of Alleo Labs and has received a research grant from the Alzheimer’s Disease Data Initiative as a William H. Gates Sr. fellow (no grant number). Yoonjeong Cha is an employee of Alleo Labs. Mohamedi N. Kagalwala is an employee of Alleo Labs.

References

1. Berdigaliyev, N.; Aljofan, M. An Overview of Drug Discovery and Development. *Future Med. Chem.* **2020**, *12*, 939–947. [CrossRef] [PubMed]
2. Wouters, O.J.; McKee, M.; Luyten, J. Estimated Research and Development Investment Needed to Bring a New Medicine to Market, 2009–2018. *JAMA* **2020**, *323*, 844–853. [CrossRef] [PubMed]
3. Schneider, G. Automating Drug Discovery. *Nat. Rev. Drug Discov.* **2018**, *17*, 97–113. [CrossRef] [PubMed]
4. Sun, D.; Gao, W.; Hu, H.; Zhou, S. Why 90% of Clinical Drug Development Fails and How to Improve It? *Acta Pharm. Sin. B* **2022**, *12*, 3049–3062. [CrossRef] [PubMed]
5. Wong, C.H.; Siah, K.W.; Lo, A.W. Estimation of Clinical Trial Success Rates and Related Parameters. *Biostatistics* **2019**, *20*, 273–286. [CrossRef]
6. Lamprey, R.N.L.; Chaulagain, B.; Trivedi, R.; Gothwal, A.; Layek, B.; Singh, J. A Review of the Common Neurodegenerative Disorders: Current Therapeutic Approaches and the Potential Role of Nanotherapeutics. *Int. J. Mol. Sci.* **2022**, *23*, 1851. [CrossRef]
7. Filippi, M.; Bar-Or, A.; Piehl, F.; Preziosa, P.; Solari, A.; Vukusic, S.; Rocca, M.A. Multiple Sclerosis. *Nat. Rev. Dis. Primers* **2018**, *4*, 43. [CrossRef]
8. Alzheimer’s Disease Fact Sheet. Available online: <https://www.nia.nih.gov/health/alzheimers-and-dementia/alzheimers-disease-fact-sheet> (accessed on 27 December 2023).
9. McKhann, G.; Drachman, D.; Folstein, M.; Katzman, R.; Price, D.; Stadlan, E.M. Clinical Diagnosis of Alzheimer’s Disease: Report of the NINCDS-ADRDA Work Group under the Auspices of Department of Health and Human Services Task Force on Alzheimer’s Disease. *Neurology* **1984**, *34*, 939–944. [CrossRef]
10. Epperly, T.; Dunay, M.A.; Boice, J.L. Alzheimer Disease: Pharmacologic and Nonpharmacologic Therapies for Cognitive and Functional Symptoms. *Am. Fam. Physician* **2017**, *95*, 771–778.
11. What Is ALS?—Amyotrophic Lateral Sclerosis | The ALS Association. Available online: <https://www.als.org/understanding-als/what-is-als#:~:text=There%20are%20currently%20seven%20drugs,are%20ongoing%20around%20the%20world> (accessed on 22 December 2023).
12. Jiang, J.; Wang, Y.; Deng, M. New Developments and Opportunities in Drugs Being Trialed for Amyotrophic Lateral Sclerosis from 2020 to 2022. *Front. Pharmacol.* **2022**, *13*, 1054006. [CrossRef]
13. Wong, C.; Stavrou, M.; Elliott, E.; Gregory, J.M.; Leigh, N.; Pinto, A.A.; Williams, T.L.; Chataway, J.; Swingler, R.; Parmar, M.K.B.; et al. Clinical Trials in Amyotrophic Lateral Sclerosis: A Systematic Review and Perspective. *Brain Commun.* **2021**, *3*, fcab242. [CrossRef]
14. Jellinger, K.A. Recent Update on the Heterogeneity of the Alzheimer’s Disease Spectrum. *J. Neural Transm.* **2022**, *129*, 1–24. [CrossRef]
15. Avelar-Pereira, B.; Belloy, M.E.; O’Hara, R.; Hosseini, S.M.H.; Alzheimer’s Disease Neuroimaging Initiative. Decoding the Heterogeneity of Alzheimer’s Disease Diagnosis and Progression Using Multilayer Networks. *Mol. Psychiatry* **2023**, *28*, 2423–2432. [CrossRef]
16. Tijms, B.M.; Vromen, E.M.; Mjaavatten, O.; Holstege, H.; Reus, L.M.; van der Lee, S.; Wesenhagen, K.E.; Lorenzini, L.; Vermunt, L.; Venkatraghavan, V.; et al. Large-Scale Cerebrospinal Fluid Proteomic Analysis in Alzheimer’s Disease Patients Reveals Five Molecular Subtypes with Distinct Genetic Risk Profiles. *medRxiv* **2023**. [CrossRef]
17. Neff, R.A.; Wang, M.; Vatansever, S.; Guo, L.; Ming, C.; Wang, Q.; Wang, E.; Horgusluoglu-Moloch, E.; Song, W.-M.; Li, A.; et al. Molecular Subtyping of Alzheimer’s Disease Using RNA Sequencing Data Reveals Novel Mechanisms and Targets. *Sci. Adv.* **2021**, *7*, eabb5398. [CrossRef] [PubMed]
18. Ferreira, D.; Pereira, J.B.; Volpe, G.; Westman, E. Subtypes of Alzheimer’s Disease Display Distinct Network Abnormalities Extending Beyond Their Pattern of Brain Atrophy. *Front. Neurol.* **2019**, *10*, 524. [CrossRef] [PubMed]
19. Goyal, N.A.; Berry, J.D.; Windebank, A.; Staff, N.P.; Maragakis, N.J.; van den Berg, L.H.; Genge, A.; Miller, R.; Baloh, R.H.; Kern, R.; et al. Addressing Heterogeneity in Amyotrophic Lateral Sclerosis CLINICAL TRIALS. *Muscle Nerve* **2020**, *62*, 156–166. [CrossRef] [PubMed]
20. Beghi, E.; Mennini, T.; Bendotti, C.; Bigini, P.; Logroscino, G.; Chiò, A.; Hardiman, O.; Mitchell, D.; Swingler, R.; Traynor, B.J.; et al. The Heterogeneity of Amyotrophic Lateral Sclerosis: A Possible Explanation of Treatment Failure. *Curr. Med. Chem.* **2007**, *14*, 3185–3200. [CrossRef] [PubMed]

21. Eshima, J.; O'Connor, S.A.; Marschall, E.; Bowser, R.; NYGC ALS Consortium; Plaisier, C.L.; Smith, B.S. Molecular Subtypes of ALS Are Associated with Differences in Patient Prognosis. *Nat. Commun.* **2023**, *14*, 95. [[CrossRef](#)] [[PubMed](#)]
22. Greenland, J.C.; Williams-Gray, C.H.; Barker, R.A. The Clinical Heterogeneity of Parkinson's Disease and Its Therapeutic Implications. *Eur. J. Neurosci.* **2019**, *49*, 328–338. [[CrossRef](#)] [[PubMed](#)]
23. Berg, D.; Borghammer, P.; Fereshtehnejad, S.-M.; Heinzel, S.; Horsager, J.; Schaeffer, E.; Postuma, R.B. Prodromal Parkinson Disease Subtypes—Key to Understanding Heterogeneity. *Nat. Rev. Neurol.* **2021**, *17*, 349–361. [[CrossRef](#)] [[PubMed](#)]
24. Schalkamp, A.-K.; Rahman, N.; Monzón-Sandoval, J.; Sandor, C. Deep Phenotyping for Precision Medicine in Parkinson's Disease. *Dis. Model. Mech.* **2022**, *15*, dmm049376. [[CrossRef](#)] [[PubMed](#)]
25. Wattamwar, P.R.; Mathuranath, P.S. An Overview of Biomarkers in Alzheimer's Disease. *Ann. Indian Acad. Neurol.* **2010**, *13*, S116–S123. [[CrossRef](#)] [[PubMed](#)]
26. Xiao, X.; Liu, H.; Liu, X.; Zhang, W.; Zhang, S.; Jiao, B. APP, PSEN1, and PSEN2 Variants in Alzheimer's Disease: Systematic Re-Evaluation According to ACMG Guidelines. *Front. Aging Neurosci.* **2021**, *13*, 695808. [[CrossRef](#)] [[PubMed](#)]
27. Aasly, J.O. Long-Term Outcomes of Genetic Parkinson's Disease. *J. Mov. Disord.* **2020**, *13*, 81–96. [[CrossRef](#)]
28. Sirkis, D.W.; Bonham, L.W.; Johnson, T.P.; La Joie, R.; Yokoyama, J.S. Dissecting the Clinical Heterogeneity of Early-Onset Alzheimer's Disease. *Mol. Psychiatry* **2022**, *27*, 2674–2688. [[CrossRef](#)]
29. Jack, C.R.; Bernstein, M.A.; Fox, N.C.; Thompson, P.; Alexander, G.; Harvey, D.; Borowski, B.; Britson, P.J.; Whitwell, J.L.; Ward, C.; et al. The Alzheimer's Disease Neuroimaging Initiative (ADNI): MRI Methods. *J. Magn. Reson. Imaging* **2008**, *27*, 685–691. [[CrossRef](#)]
30. Toga, A.W.; Phatak, M.; Pappas, I.; Thompson, S.; McHugh, C.P.; Clement, M.H.S.; Bauermeister, S.; Maruyama, T.; Gallacher, J. The Pursuit of Approaches to Federate Data to Accelerate Alzheimer's Disease and Related Dementia Research: GAAIN, DPUK, and ADDI. *Front. Neuroinform.* **2023**, *17*, 1175689. [[CrossRef](#)]
31. Bennett, D.A.; Buchman, A.S.; Boyle, P.A.; Barnes, L.L.; Wilson, R.S.; Schneider, J.A. Religious Orders Study and Rush Memory and Aging Project. *J. Alzheimers Dis.* **2018**, *64*, S161–S189. [[CrossRef](#)] [[PubMed](#)]
32. Hodes, R.J.; Buckholtz, N. Accelerating Medicines Partnership: Alzheimer's Disease (AMP-AD) Knowledge Portal Aids Alzheimer's Drug Discovery through Open Data Sharing. *Expert. Opin. Ther. Targets* **2016**, *20*, 389–391. [[CrossRef](#)] [[PubMed](#)]
33. Marek, K.; Chowdhury, S.; Siderowf, A.; Lasch, S.; Coffey, C.S.; Caspell-Garcia, C.; Simuni, T.; Jennings, D.; Tanner, C.M.; Trojanowski, J.Q.; et al. The Parkinson's Progression Markers Initiative (PPMI)—Establishing a PD Biomarker Cohort. *Ann. Clin. Transl. Neurol.* **2018**, *5*, 1460–1477. [[CrossRef](#)]
34. Baxi, E.G.; Thompson, T.; Li, J.; Kaye, J.A.; Lim, R.G.; Wu, J.; Ramamoorthy, D.; Lima, L.; Vaibhav, V.; Matlock, A.; et al. Answer ALS, a Large-Scale Resource for Sporadic and Familial ALS Combining Clinical and Multi-Omics Data from Induced Pluripotent Cell Lines. *Nat. Neurosci.* **2022**, *25*, 226–237. [[CrossRef](#)]
35. Target ALS. Available online: <https://www.targetals.org/resource/genomic-datasets/> (accessed on 26 December 2023).
36. Malhotra, R.; Seth, I.; Lehnert, E.; Zhao, J.; Kaushik, G.; Williams, E.H.; Sethi, A.; Davis-Dusenbery, B.N. Using the Seven Bridges Cancer Genomics Cloud to Access and Analyze Petabytes of Cancer Data. *Curr. Protoc. Bioinform.* **2017**, *60*, 11–16. [[CrossRef](#)] [[PubMed](#)]
37. Guo, L.; Kong, D.; Liu, J.; Zhan, L.; Luo, L.; Zheng, W.; Zheng, Q.; Chen, C.; Sun, S. Breast Cancer Heterogeneity and Its Implication in Personalized Precision Therapy. *Exp. Hematol. Oncol.* **2023**, *12*, 3. [[CrossRef](#)] [[PubMed](#)]
38. Arakelyan, A.; Melkonyan, A.; Hakobyan, S.; Boyarskih, U.; Simonyan, A.; Nersisyan, L.; Nikoghosyan, M.; Filipenko, M.; Binder, H. Transcriptome Patterns of BRCA1- and BRCA2- Mutated Breast and Ovarian Cancers. *Int. J. Mol. Sci.* **2021**, *22*, 1266. [[CrossRef](#)] [[PubMed](#)]
39. Petersen, I. The Morphological and Molecular Diagnosis of Lung Cancer. *Dtsch. Ärzteblatt Int.* **2011**, *108*, 525–531. [[CrossRef](#)] [[PubMed](#)]
40. Fauteux, F.; Surendra, A.; McComb, S.; Pan, Y.; Hill, J.J. Identification of Transcriptional Subtypes in Lung Adenocarcinoma and Squamous Cell Carcinoma through Integrative Analysis of Microarray and RNA Sequencing Data. *Sci. Rep.* **2021**, *11*, 8709. [[CrossRef](#)] [[PubMed](#)]
41. Søkilde, R.; Persson, H.; Ehinger, A.; Pirona, A.C.; Fernö, M.; Hegardt, C.; Larsson, C.; Loman, N.; Malmberg, M.; Rydén, L.; et al. Refinement of Breast Cancer Molecular Classification by miRNA Expression Profiles. *BMC Genom.* **2019**, *20*, 503. [[CrossRef](#)] [[PubMed](#)]
42. Wilkerson, M.D.; Hayes, D.N. ConsensusClusterPlus: A Class Discovery Tool with Confidence Assessments and Item Tracking. *Bioinformatics* **2010**, *26*, 1572–1573. [[CrossRef](#)]
43. Verhaak, R.G.W.; Hoadley, K.A.; Purdom, E.; Wang, V.; Qi, Y.; Wilkerson, M.D.; Miller, C.R.; Ding, L.; Golub, T.; Mesirov, J.P.; et al. Integrated Genomic Analysis Identifies Clinically Relevant Subtypes of Glioblastoma Characterized by Abnormalities in PDGFRA, IDH1, EGFR, and NF1. *Cancer Cell* **2010**, *17*, 98–110. [[CrossRef](#)]
44. Hayes, D.N.; Monti, S.; Parmigiani, G.; Gilks, C.B.; Naoki, K.; Bhattacharjee, A.; Socinski, M.A.; Perou, C.; Meyerson, M. Gene Expression Profiling Reveals Reproducible Human Lung Adenocarcinoma Subtypes in Multiple Independent Patient Cohorts. *J. Clin. Oncol.* **2006**, *24*, 5079–5090. [[CrossRef](#)]
45. Tam, O.H.; Rozhkov, N.V.; Shaw, R.; Kim, D.; Hubbard, I.; Fennessey, S.; Propp, N.; Fagegaltier, D.; Harris, B.T.; Ostrow, L.W.; et al. Postmortem Cortex Samples Identify Distinct Molecular Subtypes of ALS: Retrotransposon Activation, Oxidative Stress, and Activated Glia. *Cell Rep.* **2019**, *29*, 1164–1177. [[CrossRef](#)]

46. Abe, K.; Ohkubo, T.; Yokota, T. TDP-43 in the Skin of Amyotrophic Lateral Sclerosis Patients. *J. Med. Dent. Sci.* **2017**, *64*, 9–17. [[CrossRef](#)]
47. Rao, M.; McDuffie, E.; Sachs, C. Artificial Intelligence/Machine Learning-Driven Small Molecule Repurposing via Off-Target Prediction and Transcriptomics. *Toxics* **2023**, *11*, 875. [[CrossRef](#)]
48. Eisenstein, M. Active Machine Learning Helps Drug Hunters Tackle Biology. *Nat. Biotechnol.* **2020**, *38*, 512–514. [[CrossRef](#)]
49. Schneider, M.; Belsom, A.; Rappsilber, J. Protein Tertiary Structure by Crosslinking/Mass Spectrometry. *Trends Biochem. Sci.* **2018**, *43*, 157–169. [[CrossRef](#)]
50. Lill, M.A.; Danielson, M.L. Computer-Aided Drug Design Platform Using PyMOL. *J. Comput. Aided Mol. Des.* **2011**, *25*, 13–19. [[CrossRef](#)]
51. Life Science: SiteMap. Available online: <https://newsite.schrodinger.com/platform/products/sitemap/> (accessed on 19 December 2023).
52. Le Guilloux, V.; Schmidtke, P.; Tuffery, P. Fpocket: An Open Source Platform for Ligand Pocket Detection. *BMC Bioinform.* **2009**, *10*, 168. [[CrossRef](#)] [[PubMed](#)]
53. Volkamer, A.; Kuhn, D.; Rippmann, F.; Rarey, M. DoGSiteScorer: A Web Server for Automatic Binding Site Prediction, Analysis and Druggability Assessment. *Bioinformatics* **2012**, *28*, 2074–2075. [[CrossRef](#)] [[PubMed](#)]
54. Laurie, A.T.; Jackson, R.M. Q-SiteFinder: An energy-based method for the prediction of protein-ligand binding sites. *Bioinformatics* **2005**, *21*, 1908–1916. [[CrossRef](#)] [[PubMed](#)]
55. Di Pietro, O.; Juárez-Jiménez, J.; Muñoz-Torrero, D.; Loughton, C.A.; Luque, F.J. Unveiling a Novel Transient Druggable Pocket in BACE-1 through Molecular Simulations: Conformational Analysis and Binding Mode of Multisite Inhibitors. *PLoS ONE* **2017**, *12*, e0177683. [[CrossRef](#)]
56. De Luca, L.; Vittorio, S.; Peña-Díaz, S.; Pitasi, G.; Fornit-Suñé, M.; Bucolo, F.; Ventura, S.; Gitto, R. Ligand-Based Discovery of a Small Molecule as Inhibitor of α -Synuclein Amyloid Formation. *Int. J. Mol. Sci.* **2022**, *23*, 14844. [[CrossRef](#)]
57. Silva, G.M.; Borges, R.S.; Santos, K.L.B.; Federico, L.B.; Francischini, I.A.G.; Gomes, S.Q.; Barcelos, M.P.; Silva, R.C.; Santos, C.B.R.; Silva, C.H.T.P. Revisiting the Proposition of Binding Pockets and Bioactive Poses for GSK-3 β Allosteric Modulators Addressed to Neurodegenerative Diseases. *Int. J. Mol. Sci.* **2021**, *22*, 8252. [[CrossRef](#)]
58. Jiménez, J.; Doerr, S.; Martínez-Rosell, G.; Rose, A.S.; De Fabritiis, G. DeepSite: Protein-Binding Site Predictor Using 3D-Convolutional Neural Networks. *Bioinformatics* **2017**, *33*, 3036–3042. [[CrossRef](#)]
59. Desaphy, J.; Bret, G.; Rognan, D.; Kellenberger, E. Sc-PDB: A 3D-Database of Ligandable Binding Sites—10 Years On. *Nucleic Acids Res.* **2015**, *43*, D399–D404. [[CrossRef](#)]
60. Casiraghi, A.; Longhena, F.; Faustini, G.; Ribaudo, G.; Suigo, L.; Camacho-Hernandez, G.A.; Bono, F.; Brembati, V.; Newman, A.H.; Gianoncelli, A.; et al. Methylphenidate Analogues as a New Class of Potential Disease-Modifying Agents for Parkinson's Disease: Evidence from Cell Models and Alpha-Synuclein Transgenic Mice. *Pharmaceutics* **2022**, *14*, 1595. [[CrossRef](#)] [[PubMed](#)]
61. Longhena, F.; Faustini, G.; Varanita, T.; Zaltieri, M.; Porrini, V.; Tessari, I.; Poliani, P.L.; Missale, C.; Borroni, B.; Padovani, A.; et al. Synapsin III Is a Key Component of A-synuclein Fibrils in Lewy Bodies of PD Brains. *Brain Pathol.* **2018**, *28*, 875–888. [[CrossRef](#)] [[PubMed](#)]
62. Szymczyzna, B.R.; Taurog, R.E.; Young, M.J.; Snyder, J.C.; Johnson, J.E.; Williamson, J.R. Synergy of NMR, Computation, and X-Ray Crystallography for Structural Biology. *Structure* **2009**, *17*, 499–507. [[CrossRef](#)] [[PubMed](#)]
63. Baek, M.; DiMaio, F.; Anishchenko, I.; Dauparas, J.; Ovchinnikov, S.; Lee, G.R.; Wang, J.; Cong, Q.; Kinch, L.N.; Schaeffer, R.D.; et al. Accurate Prediction of Protein Structures and Interactions Using a Three-Track Neural Network. *Science* **2021**, *373*, 871–876. [[CrossRef](#)] [[PubMed](#)]
64. Zhang, Y. I-TASSER Server for Protein 3D Structure Prediction. *BMC Bioinform.* **2008**, *9*, 40. [[CrossRef](#)] [[PubMed](#)]
65. Jumper, J.; Evans, R.; Pritzel, A.; Green, T.; Figurnov, M.; Ronneberger, O.; Tunyasuvunakool, K.; Bates, R.; Židek, A.; Potapenko, A.; et al. Highly Accurate Protein Structure Prediction with AlphaFold. *Nature* **2021**, *596*, 583–589. [[CrossRef](#)]
66. Xu, D.; Zhang, Y. Ab Initio Protein Structure Assembly Using Continuous Structure Fragments and Optimized Knowledge-Based Force Field. *Proteins* **2012**, *80*, 1715–1735. [[CrossRef](#)]
67. Pandey, M.; Fernandez, M.; Gentile, F.; Isayev, O.; Tropsha, A.; Stern, A.C.; Cherkasov, A. The Transformational Role of GPU Computing and Deep Learning in Drug Discovery. *Nat. Mach. Intell.* **2022**, *4*, 211–221. [[CrossRef](#)]
68. Unke, O.T.; Chmiela, S.; Sauceda, H.E.; Gastegger, M.; Poltavsky, I.; Schütt, K.T.; Tkatchenko, A.; Müller, K.-R. Machine Learning Force Fields. *Chem. Rev.* **2021**, *121*, 10142–10186. [[CrossRef](#)]
69. Quinn, P.M.J.; Moreira, P.I.; Ambrósio, A.F.; Alves, C.H. PINK1/PARKIN Signalling in Neurodegeneration and Neuroinflammation. *Acta Neuropathol. Commun.* **2020**, *8*, 189. [[CrossRef](#)]
70. Kakade, P.; Ojha, H.; Raimi, O.G.; Shaw, A.; Waddell, A.D.; Ault, J.R.; Burel, S.; Brockmann, K.; Kumar, A.; Ahangar, M.S.; et al. Mapping of a N-Terminal α -Helix Domain Required for Human PINK1 Stabilization, Serine228 Autophosphorylation and Activation in Cells. *Open Biol.* **2022**, *12*, 210264. [[CrossRef](#)]
71. Misra, A.; Chakrabarti, S.S.; Gambhir, I.S. New Genetic Players in Late-Onset Alzheimer's Disease: Findings of Genome-Wide Association Studies. *Indian J. Med. Res.* **2018**, *148*, 135–144. [[CrossRef](#)] [[PubMed](#)]
72. Zhao, J.; Liu, X.; Xia, W.; Zhang, Y.; Wang, C. Targeting Amyloidogenic Processing of APP in Alzheimer's Disease. *Front. Mol. Neurosci.* **2020**, *13*, 137. [[CrossRef](#)] [[PubMed](#)]

73. Yamazaki, Y.; Painter, M.M.; Bu, G.; Kanekiyo, T. Apolipoprotein E as a Therapeutic Target in Alzheimer's Disease: A Review of Basic Research and Clinical Evidence. *CNS Drugs* **2016**, *30*, 773–789. [CrossRef]
74. Singh, H.; Rai, V.; Nooti, S.K.; Agrawal, D.K. Novel Ligands and Modulators of Triggering Receptor Expressed on Myeloid Cells Receptor Family: 2015-2020 Updates. *Expert Opin. Ther. Pat.* **2021**, *31*, 549–561. [CrossRef]
75. Serneels, L.; Narlawar, R.; Perez-Benito, L.; Muncioy, M.; Guallar, V.; T'Syen, D.; Dewilde, M.; Bischoff, F.; Fraiponts, E.; Tresadern, G.; et al. Selective Inhibitors of the PSEN1-Gamma-Secretase Complex. *J. Biol. Chem.* **2023**, *299*, 104794. [CrossRef]
76. Zhang, Y.; Skolnick, J. TM-Align: A Protein Structure Alignment Algorithm Based on the TM-Score. *Nucleic Acids Res.* **2005**, *33*, 2302–2309. [CrossRef]
77. Efraimidis, E.; Krokidis, M.G.; Exarchos, T.P.; Lazar, T.; Vlamos, P. In Silico Structural Analysis Exploring Conformational Folding of Protein Variants in Alzheimer's Disease. *Int. J. Mol. Sci.* **2023**, *24*, 13543. [CrossRef]
78. Subramaniam, S.; Mehrotra, M.; Gupta, D. Virtual High Throughput Screening (vHTS)—A Perspective. *Bioinformatics* **2008**, *3*, 14–17. [CrossRef] [PubMed]
79. Mayr, L.M.; Bojanic, D. Novel Trends in High-Throughput Screening. *Curr. Opin. Pharmacol.* **2009**, *9*, 580–588. [CrossRef] [PubMed]
80. Kimber, T.B.; Chen, Y.; Volkamer, A. Deep Learning in Virtual Screening: Recent Applications and Developments. *Int. J. Mol. Sci.* **2021**, *22*, 4435. [CrossRef] [PubMed]
81. D'Souza, S.; Prema, K.V.; Balaji, S.; Shah, R. Deep Learning-Based Modeling of Drug–Target Interaction Prediction Incorporating Binding Site Information of Proteins. *Interdiscip. Sci.* **2023**, *15*, 306–315. [CrossRef] [PubMed]
82. Bragina, M.E.; Daina, A.; Perez, M.A.S.; Michielin, O.; Zoete, V. The SwissSimilarity 2021 Web Tool: Novel Chemical Libraries and Additional Methods for an Enhanced Ligand-Based Virtual Screening Experience. *Int. J. Mol. Sci.* **2022**, *23*, 811. [CrossRef] [PubMed]
83. Peña-Díaz, S.; Pujols, J.; Ventura, S. Small Molecules to Prevent the Neurodegeneration Caused by α -Synuclein Aggregation. *Neural Regen. Res.* **2020**, *15*, 2260–2261. [CrossRef]
84. Daina, A.; Zoete, V. Application of the SwissDrugDesign Online Resources in Virtual Screening. *Int. J. Mol. Sci.* **2019**, *20*, 4612. [CrossRef]
85. Wang, X.; Zhong, Y.; Ding, M. Repositioning Drugs to the Mitochondrial Fusion Protein 2 by Three-Tunnel Deep Neural Network for Alzheimer's Disease. *Front. Genet.* **2021**, *12*, 638330. [CrossRef]
86. Engelhart, E.A.; Hoppins, S. A Catalytic Domain Variant of Mitofusin Requiring a Wildtype Paralog for Function Uncouples Mitochondrial Outer-Membrane Tethering and Fusion. *J. Biol. Chem.* **2019**, *294*, 8001–8014. [CrossRef] [PubMed]
87. Stepniewska-Dziubinska, M.M.; Zielenkiewicz, P.; Siedlecki, P. Improving Detection of Protein-Ligand Binding Sites with 3D Segmentation. *Sci. Rep.* **2020**, *10*, 5035. [CrossRef] [PubMed]
88. Pitt, D.; Werner, P.; Raine, C.S. Glutamate Excitotoxicity in a Model of Multiple Sclerosis. *Nat. Med.* **2000**, *6*, 67–70. [CrossRef] [PubMed]
89. Zhai, D.; Yan, S.; Samsom, J.; Wang, L.; Su, P.; Jiang, A.; Zhang, H.; Jia, Z.; Wallach, I.; Heifets, A.; et al. Small-Molecule Targeting AMPA-Mediated Excitotoxicity Has Therapeutic Effects in Mouse Models for Multiple Sclerosis. *Sci. Adv.* **2023**, *9*, ead6187. [CrossRef]
90. Salpietro, V.; Dixon, C.L.; Guo, H.; Bello, O.D.; Vandrovicova, J.; Efthymiou, S.; Maroofian, R.; Heimer, G.; Burglen, L.; Valence, S.; et al. AMPA Receptor GluA2 Subunit Defects Are a Cause of Neurodevelopmental Disorders. *Nat. Commun.* **2019**, *10*, 3094. [CrossRef]
91. Wyatt, P.G.; Gilbert, I.H.; Read, K.D.; Fairlamb, A.H. Target Validation: Linking Target and Chemical Properties to Desired Product Profile. *Curr. Top. Med. Chem.* **2011**, *11*, 1275–1283. [CrossRef] [PubMed]
92. Hughes, J.; Rees, S.; Kalindjian, S.; Philpott, K. Principles of Early Drug Discovery. *Br. J. Pharmacol.* **2011**, *162*, 1239–1249. [CrossRef]
93. Kumar, R.; Sharma, A.; Alexiou, A.; Bilgrami, A.L.; Kamal, M.A.; Ashraf, G.M. DeePred-BBB: A Blood Brain Barrier Permeability Prediction Model With Improved Accuracy. *Front. Neurosci.* **2022**, *16*, 858126. [CrossRef]
94. Li, J.; Jiang, X. Mol-BERT: An Effective Molecular Representation with BERT for Molecular Property Prediction. *Wirel. Commun. Mob. Comput.* **2021**, *2021*, 7181815. [CrossRef]
95. Verge Genomics. Available online: <https://www.vergegenomics.com> (accessed on 26 December 2023).
96. Corporation, I. Immuneering Enters Collaboration with Astex Pharmaceuticals to Identify Novel Therapeutic Targets in Central Nervous System Disorder Using Disease Cancelling Technology. Available online: <https://www.globenewswire.com/en/news-release/2020/06/25/2053543/0/en/Immuneering-Enters-Collaboration-with-Astex-Pharmaceuticals-to-Identify-Novel-Therapeutic-Targets-in-Central-Nervous-System-Disorder-Using-Disease-Cancelling-Technology.html> (accessed on 26 December 2023).
97. Making Medicines Differently. Available online: <https://www.insitro.com/> (accessed on 26 December 2023).
98. Evotec Reaches First Milestone and Receives Success Payment in Alzheimer's Disease Collaboration with Takeda—Evotec. Available online: <https://www.evotec.com/en/investor-relations/news/p/evotec-reaches-first-milestone-and-receives-success-payment-in-alzheimers-disease-collaboration-with-takeda-4533> (accessed on 26 December 2023).
99. Welcome to InveniAI. Available online: <https://www.inveniai.com/> (accessed on 26 December 2023).

100. Recursion Recursion Announces Transformational Collaboration with Roche and Genentech in Neuroscience and Oncology, Advancing Novel Medicines to Patients Using Machine Learning and High Content Screening Methods at Scale to Map Complex Biology. Available online: <https://www.prnewswire.com/news-releases/recursion-announces-transformational-collaboration-with-roche-and-genentech-in-neuroscience-and-oncology-advancing-novel-medicines-to-patients-using-machine-learning-and-high-content-screening-methods-at-scale-to-map-complex-biol-301438560.html> (accessed on 26 December 2023).
101. Hale, C. JPM23: Schrödinger Expands Neuroscience Work with BMS, Otsuka Partnerships. Available online: <https://www.fiercebiotech.com/medtech/jpm23-schrodinger-expands-neuroscience-work-bms-otsuka-partnerships> (accessed on 26 December 2023).
102. Our Team—WaveBreak. Available online: <https://wavebreaktx.com/our-team/> (accessed on 26 December 2023).
103. End-to-End Drug Discovery. Available online: <https://www.benevolent.com/benevolent-platform/end-end-drug-discovery/> (accessed on 26 December 2023).
104. Merck KGaA Taps BenevolentAI, Exscientia for AI Drug Discovery. Available online: <https://www.fiercebiotech.com/biotech/merck-kgaa-doubles-ai-partners-tapping-benevolentai-and-exscientia-drug-discovery-push> (accessed on 26 December 2023).
105. Vincere Biosciences Inc. Available online: <https://vincerebio.com/> (accessed on 26 December 2023).
106. Alleo Labs. Available online: <https://www.alleolabs.com> (accessed on 27 December 2023).
107. AbbVie Partners with BigHat to Develop Antibody Therapies. Available online: <https://www.pharmaceutical-technology.com/news/abbvie-partners-with-bighat-to-develop-antibody-therapies> (accessed on 26 December 2023).
108. Technology. Available online: <https://www.bighatbio.com/technology> (accessed on 26 December 2023).
109. Verge Genomics Announces Positive Safety and Tolerability Data from the Phase 1 Clinical Trial of VRG50635, a Potential Best-in-Class Therapeutic for All Forms of ALS. Available online: <https://www.biospace.com/article/verge-genomics-announces-positive-safety-and-tolerability-data-from-the-phase-1-clinical-trial-of-vrg50635-a-potential-best-in-class-therapeutic-for-all-forms-of-als/> (accessed on 26 December 2023).
110. NeuBio. Available online: <https://www.neu.bio> (accessed on 26 December 2023).
111. Home. Available online: <https://www.perceiv.ai/> (accessed on 26 December 2023).
112. Rune Labs. Available online: <https://www.runelabs.io/> (accessed on 26 December 2023).
113. Home—LinusBio. Available online: <https://linusbio.com/> (accessed on 26 December 2023).
114. Inception Program for Startups | NVIDIA. Available online: <https://www.nvidia.com/en-us/startups/> (accessed on 26 December 2023).

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.