ACCSCIENCE PUBLISHING

Artificial Intelligence in Health

REVIEW ARTICLE

Application and design of artificial intelligence in computer-aided drug design for ischemic stroke

Shutong Pang, Yuehui Liao, Yu Hu, Yu Chen, Xiaobo Lai*, and Panfei Li*

School of Medical Technology and Information Engineering, Zhejiang Chinese Medical University, Hangzhou, China

Abstract

The development of drugs for ischemic stroke has long been plagued by the dilemma of high investment, high attrition, and low translational success rates. This narrative review summarizes the application of artificial intelligence (AI) in ischemic stroke drug design, specifically encompassing intelligent target mining, the generation and optimization of lead compounds, early prediction of pharmacokinetics/toxicity (ADMET), and the optimization of clinical trial protocols. Although AI-aided drug design has progressed from a theoretical concept to practical validation, significant challenges remain in data quality, model interpretability, and clinical translation. However, with ongoing efforts such as the construction of high-quality multimodal databases and the advancement of explainable AI models, artificial intelligence holds considerable promise for continuously driving the discovery of novel therapeutics for stroke treatment.

Keywords: Artificial intelligence; Ischemic stroke; Computer-aided drug design; Drug discovery; Machine learning

***Corresponding author:**
Xiaobo Lai
(dmia_lab@zcmu.edu.cn)
Panfei Li
(lipanfei@zcmu.edu.cn)

Citation: Pang S, Liao Y, Hu Y, Chen Y, Lai X, Li P. Application and design of artificial intelligence in computer-aided drug design for ischemic stroke. *Artif Intell Health*. doi: 10.36922/AIH026170038

Received: April 23, 2026

Revised: May 21, 2026

Accepted: May 27, 2026

Published online: June 16, 2026

Copyright: © 2026 Author(s). This is an Open-Access article distributed under the terms of the Creative Commons Attribution License, permitting distribution, and reproduction in any medium, provided the original work is properly cited.

Publisher's Note: AccScience Publishing remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

1. Introduction

Stroke is an acute focal neurological deficit syndrome and represents the third leading cause of death and the primary cause of long-term disability globally. Common clinical manifestations encompass hemiplegia, dysarthria, sensory disturbances, aphasia, and visual impairments.[1] Ischemic stroke, which results from acute arterial occlusion, constitutes 60% to 70% of all stroke cases worldwide. It is characterized by high incidence, disability, mortality, and recurrence rates, with this proportion remaining relatively consistent across different geographic regions.[2] The disease imposes a substantial socioeconomic burden, a trend projected to be further exacerbated by the aging global population.[3,4]

The clinical management of ischemic stroke has long centered on the two primary strategies of reperfusion therapy and neuroprotection.[4] In the acute phase, recombinant tissue plasminogen activator (rtPA) stands as the sole pharmacological agent approved by the U.S. Food and Drug Administration for achieving recanalization via thrombolysis. Nevertheless, its narrow therapeutic window (approximately 4.5 h) coupled with the inherent risk of hemorrhagic transformation renders fewer than 10% of patients eligible for this treatment, consequently depriving the vast majority of its potential benefits.[5,6] Meanwhile, the field of neuroprotection has amassed a substantial body of preclinical

trial design over the past three decades. This work has elucidated key pathological mechanisms underpinning ischemic injury, including oxidative stress, excitotoxicity, and inflammation. Interventions targeting these pathways have demonstrated efficacy in mitigating neuronal damage across various experimental models.[7] Nevertheless, the translation of these mechanistically sound concepts into neuroprotective agents with demonstrable clinical efficacy remains a formidable and persistent challenge.[8]

Traditional drug discovery follows a linear path encompassing target identification and validation, lead compound generation and optimization, preclinical candidates, and subsequent clinical trials. This process is not only capital-intensive but also carries a significant risk of late-stage failure.[9] Currently, drug development for ischemic stroke faces two prominent challenges that urgently need to be addressed. The first includes the narrow therapeutic window of thrombolytic therapy and its high bleeding risk, which prevent the vast majority of patients from receiving effective treatment.[5] The second is that numerous neuroprotective agents showing good efficacy in preclinical models have repeatedly failed in clinical trials. The underlying reason is that the pathophysiology of stroke involves the interplay of multiple pathways such as oxidative stress, inflammation, and apoptosis; the traditional linear, single-target, single-drug development paradigm is inadequate to cope with such complexity.[10] The emergence of computer-aided drug design (CADD) technology has addressed these challenges by enabling the pre-screening and evaluation of compounds at a computational level. This approach effectively reduces the number of molecules that must advance to experimental validation, thereby concentrating resources on the most promising candidates and significantly alleviating the scale-, time-, and cost-related pressures inherent in purely experimental methods. Driven by the exponential growth in computing power and the increasing sophistication of computational chemistry and biology methods, CADD is now deeply integrated into virtually all stages of drug discovery and development.[11] Its application accelerates the research pipeline and helps reduce associated costs and risks in both preclinical and clinical phases.[12] In recent years, driven by innovations in machine learning (ML, a technology that learns patterns from data and makes predictions through algorithms) and the rapid development of its advanced branch, deep learning (DL, which uses multi-layered neural networks to automatically extract high-level features from data), as well as the accumulation of large-scale pharmacological data, artificial intelligence (AI) has become a powerful data mining and pattern recognition engine in CADD.[13,14] AI has achieved notable success in diverse areas, including virtual screening, activity prediction, quantitative structure-activity relationship modeling and the early prediction of pharmacokinetic and toxicity properties. Although providing clear physical interpretations for AI-based models remains a challenge, their ability to deliver diverse and innovative frameworks is undeniably reshaping the paradigm of drug discovery.[9,15] This convergence offers a revolutionary toolkit for tackling the longstanding bottlenecks in developing therapeutic agents for ischemic stroke. However, the application of AI in drug discovery is not evolving in isolation. Relevant studies have indicated that positioning AI as an assistive tool rather than a replacement, improving model decision transparency, and achieving human-machine synergy are key factors determining whether AI can be successfully implemented in clinical practice.[16]

This narrative review summarizes the transformative applications of AI in CADD for ischemic stroke. It highlights AI-driven advances in key stages of drug discovery, including the intelligent mining of novel therapeutic targets, structure-based virtual screening and lead compound generation, multi-target drug design strategies, as well as clinical trial design and patient stratification. By synthesizing progress across these core domains, this article aims to elucidate how AI is breaking through traditional research and development paradigms, thereby providing innovative solutions and a powerful new toolkit to address the longstanding bottlenecks in the development of therapeutic agents for ischemic stroke. In addition, the abbreviations involved in the article are shown in Table 1.

Table 1. Symbols and abbreviations used in the article

No.	Full name	Abbreviation
1	Computer-aided drug design	CADD
2	Artificial intelligence	AI
3	Machine learning	ML
4	Deep learning	DL
5	Generative adversarial networks	GANs
6	Variational autoencoders	VAEs
7	Pharmacokinetics and toxicity	ADMET
8	Graph neural networks	GNNs
9	Random forest	RF
10	Support vector machine	SVM
11	Neural network	NN
12	Generalized linear models	GLMs
13	Large language model	LLM
14	Magnetic resonance imaging	MRI
15	Pyroptosis-related differentially expressed genes	PRDEGs
16	Lactylation-related genes	LRGs

2. Methodology

This review follows the reporting standards for narrative reviews (SANRA guidelines) and employs a systematic literature search approach. The search was conducted in the Web of Science Core Collection and PubMed databases, covering the period from January 2015 to March 2026 to comprehensively capture relevant studies. Search keywords were designed around the core themes, including "ischemic stroke," "artificial intelligence," "machine learning," "drug discovery," "computer-aided drug design," "virtual screening," "neuroprotection," and "clinical translation." These terms were combined to construct search strategies that ensure both breadth and precision.

The PICOS framework was used to define the inclusion criteria: (1) Population: drug targets or disease models related to ischemic stroke; (2) Intervention: explicit application of AI, ML, or DL techniques; (3) Comparison: traditional drug design methods or computational approaches without AI assistance (not mandatory); (4) Outcome: drug discovery efficiency, prediction accuracy, candidate drug quality, *etc.*; (5) Study type: original research and review articles that explicitly investigate the application of AI or ML techniques in ischemic stroke drug design or related pharmacological research. Studies irrelevant to the core theme, non-English publications, conference abstracts, editorials, and commentaries were excluded. The search was not restricted by publication date, journal impact factor, or author affiliations to capture a complete landscape of the field. All retrieved records were first deduplicated using EndNote. Subsequently, two independent reviewers screened the titles and abstracts, and full texts of potentially eligible studies were obtained and re-screened to minimize selection bias. After deduplication and screening, the key findings and methodologies from eligible studies were extracted, synthesized, and integrated. These formed the basis of the discussion in the respective sections of this article.

3. Application of AI across the entire pipeline of ischemic stroke drug discovery and development

AI technology, leveraging its powerful capabilities in data processing and predictive modeling, has been deeply integrated into the entire pipeline of ischemic stroke drug research and development. As illustrated in Figure 1, AI accomplishes this through the integration of multimodal data, including multi-omics, clinical imaging, and electronic health records, and by utilizing ML and DL models for in-depth mining and analysis. This integration facilitates the intelligent identification of critical intervention nodes and multi-target drug candidates, enables precise drug design and optimization, and supports

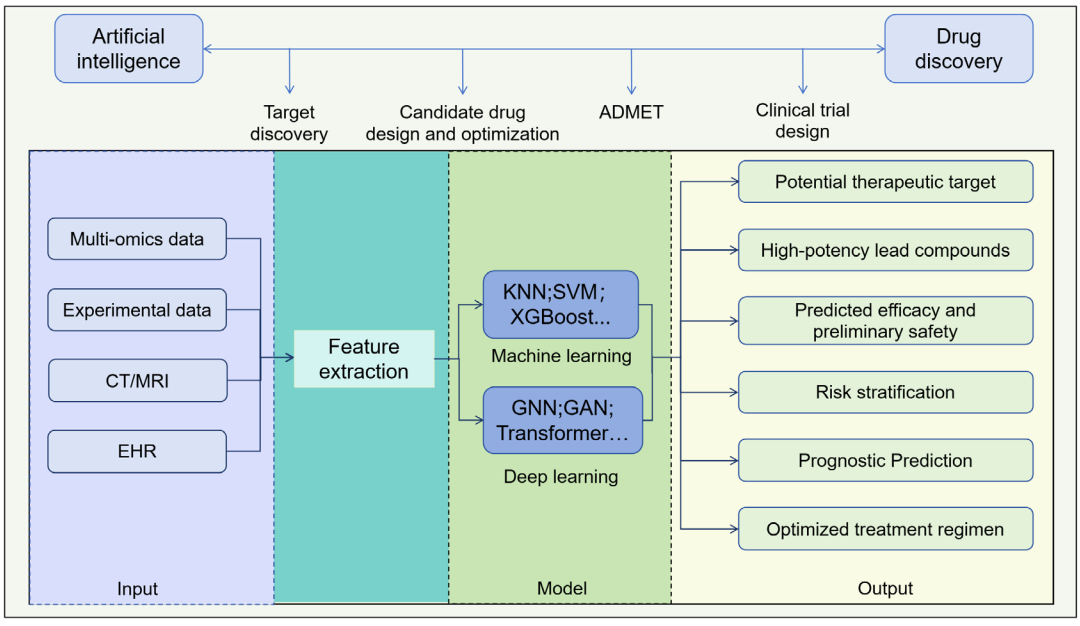


Figure 1. A workflow framework for the integrated application of AI technologies in ischemic stroke drug discovery and development
Abbreviations: ADMET: Absorption, distribution, metabolism, excretion, and toxicity; CT: Computed tomography; EHR: Electronic health record; GAN: Generative adversarial network; GNN: Graph neural network; KNN: K-nearest neighbors; MRI: Magnetic resonance imaging; SVM: Support vector machine; XGBoost: Extreme gradient boosting.

the personalized design of clinical trial protocols alongside patient stratification.[17,18] Consequently, AI is instigating transformative changes at every stage of the pipeline. To provide a structured overview of AI applications across different phases, Table 2 summarizes the typical input representations, task types, output applications, common model families, evaluation metrics, and potential challenges associated with each major stage of ischemic stroke drug discovery, offering a reference framework for the subsequent detailed discussion of AI in this field.

3.1. Drug target discovery

The identification of effective drug targets is the prerequisite and cornerstone of drug development for ischemic stroke. Its pathophysiological process involves multiple interconnected pathways, including oxidative stress, inflammatory responses, cell apoptosis, and neurovascular unit dysfunction.[19,20] This complexity subjects traditional drug discovery methods focused on single targets to severe challenges characterized by long cycles, high costs, and low success rates. The introduction of AI technology offers a new paradigm for systematically addressing this challenge. The core strength of AI lies in its ability to mine complex patterns and associations from massive, high-dimensional biomedical data, including genomic, transcriptomic, proteomic datasets and clinical literature databases, which are difficult for humans to discern intuitively.[21]

By integrating multi-omics data, AI models can construct disease-specific molecular interaction networks and utilize topological network analysis to precisely identify key nodes that are crucial for maintaining the stability of

Table 2. Summary of AI tasks and models in ischemic stroke CADD

Stage	Typical input	Task type	Output and decision use	Common model	Evaluation metrics	Common pitfalls
Target discovery	Multi-omics data, gene expression matrices, molecular interaction networks	Classification/ Regression/ Network analysis	Key gene/protein ranking; used to prioritize drug targets	RF, SVM, GNNs	AUC-ROC, precision-recall	Data heterogeneity, batch effects, incomplete functional annotations
Virtual screening	SMILES, molecular graphs, 3D conformers, protein structures	Classification/ Regression/ Ranking	Binding affinity scores, virtual screening rankings; used to narrow experimental validation	GNNs, Transformers, XGBoost	AUC-ROC, enrichment factor, BEDROC, F1-score, RMSD	Data leakage, scoring function simplifications
Molecular generation	SMILES, molecular graphs, target property constraints	Generation	New molecular structures (SMILES/molecular graphs); used to explore novel chemical space and generate leads	GANs, VAEs, Diffusion models, LSRM	Validity, Uniqueness, Novelty, FCD, Internal diversity	Size trap, training set memorization, low synthetic accessibility
ADMET prediction	SMILES, molecular graphs, molecular descriptors	Classification/ Regression	Absorption/distribution/ metabolism/excretion/ toxicity property predictions; used for early filtering of poor Pharmacokinetics/toxicity profiles	GNNs, Transformers, XGBoost, RF	AUC-ROC, F1-score, R^2, RMSE, MCC	Data imbalance, out-of-domain extrapolation, scarce BBB data, endpoint-specific
Clinical trial optimization	Electronic health records, imaging features, clinical variables	Classification/ Regression/ Clustering	Patient stratification labels, outcome risk scores, treatment response predictions; used for trial enrichment and personalized medicine	XGBoost, DL, Clustering algorithms	AUC-ROC, Calibration curves	Confounding bias, domain shift, multi-center heterogeneity, insufficient interpretability

Abbreviations: AUC-ROC: Area under the receiver operating characteristic curve; BEDROC: Boltzmann-enhanced discrimination of receiver operating characteristic; CADD: Computer-aided drug design; DL: Deep learning; FCD: Fréchet ChemNet distance; GAN: Generative adversarial network; GNN: Graph neural network; LSRM: Ligand structure–receptor matching; MCC: Matthews correlation coefficient; RF: Random forest; RMSD: Root mean square deviation; RMSE: Root mean square error; SVM: Support vector machine; VAEs: Variational autoencoders; XGBoost: Extreme gradient boosting.

the entire pathological network. This strategy enables the efficient identification of potential therapeutic targets and has been widely adopted in numerous studies that integrate bioinformatics and ML to identify core biomarkers for stroke.[22] As these identified nodes often serve as hub proteins that can intervene in multiple downstream cell death pathways, targeting them holds promise for achieving coordinated regulation of multiple pathological cascades, thereby more effectively deconstructing the entire injury network. In recent years, this concept has been successfully applied to target discovery in ischemic stroke. For instance, Peng *et al.*[17] developed a hybrid AI framework (strokeDTI) that utilizes transcriptomic sequencing data from stroke models to parse activated cell death pathways, construct gene interaction networks, and subsequently pinpoint key bottleneck genes within the network as potential targets. The strokeDTI framework then employed multiple DL architectures to predict interactions between these targets and known compounds, ultimately identifying cerdulatinib as a potential anti-stroke drug candidate. Subsequent experimental studies validated its multi-pathway protective effects.[17]

Concurrently, AI is also advancing the precise mining of targets for specific pathological mechanisms. Hei *et al.*[23] systematically analyzed pyroptosis-related differentially expressed genes in ischemic stroke patient datasets using ML algorithms. By integrating and cross-validating multiple gene expression datasets, the researchers successfully identified ten core pyroptosis-related genes crucial to the stroke pathological process, providing novel, data-driven target clues for drug development aimed specifically at inflammatory cell death.[23] Similarly, Liu *et al.*[24] utilized algorithms such as RF to identify key pyroptosis-related genes like ANXA3 and validated its role in regulating pyroptosis via the NLRC4/AIM2 axis. Furthermore, the application of AI has extended beyond traditional genomic scope to explore novel post-translational modifications such as lactylation. Through ML, hub genes like *SLC2A3* have been identified, revealing new connections between the immune microenvironment and stroke pathology.[25] Figure 2 systematically delineates the target discovery workflow spanning from multi-omics data input to candidate drug validation, thereby summarizing the core methodologies and technical stages of AI-driven target identification described in this section. The representative AI-powered target mining studies (summarized in Table 3) collectively demonstrate that integrating multi-omics data with ML approaches enables the efficient identification of core genes and precise targets associated with specific pathological mechanisms from complex disease networks.

3.2. Candidate drug design and optimization

Following the identification of potential therapeutic targets, the design and optimization of candidate drugs represent a critical bridge connecting fundamental research to clinical translation. Traditional candidate screening, which relies heavily on high-throughput experimental assays, is constrained by high costs, limited throughput, and protracted timelines. These limitations

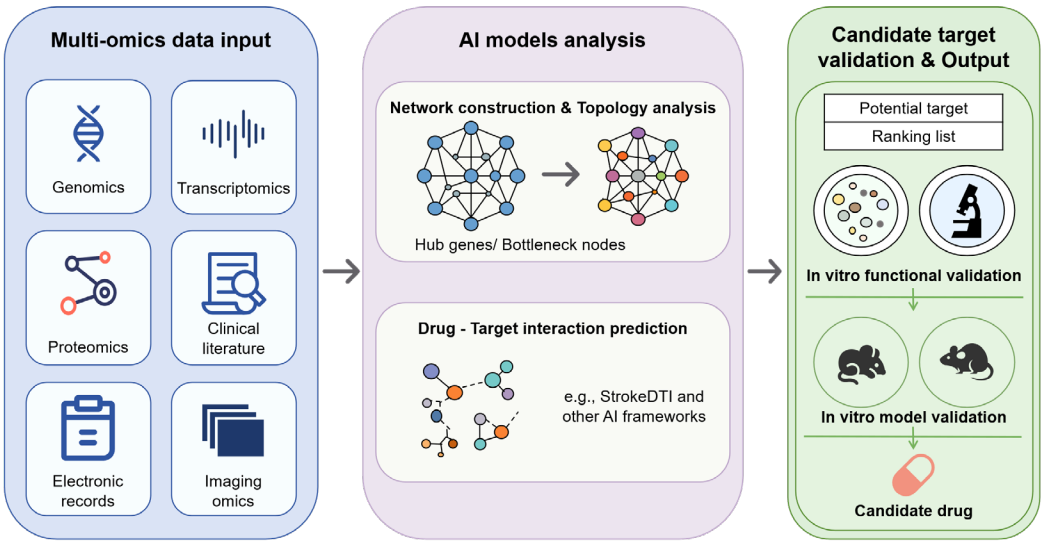


Figure 2. Flowchart of AI-driven target discovery based on multi-omics data

Table 3. AI in target discovery for ischemic stroke

No.	AI model	Application in stroke CADD	Key findings	Input data type	Validation strategy	Code/Data availability
1	StrokeDTI (DL model)	Identifying disease-specific drug-target interaction.	Identified cerdulatinib as a potential anti-stroke agent.[17]	Transcriptomic data (RNA-seq)	Experimental validation (*in vivo* and *in vitro*)	https://github.com/Moonscape/strokeDTI
2	LASSO/RF/SVM	Analyze the gene expression pattern of PRDEGs.	A set of 10 PRDEGs was identified.[23]	Gene expression datasets	Cross-validation	https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE22255https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE140275https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE16561
3	RF	Reveal the key molecular markers associated with pyroptosis in ischemic stroke.	ANXA3 as a possible pyroptosis-related gene marker in ischemic stroke.[24]	Gene expression datasets	Experimental validation (*in vivo* and *in vitro*)	Requested by contacting the first author or corresponding author.
4	RF/SVM/NN/GLMs	Analyze gene expression omnibus datasets.	Identified 6 hub LRGs associated with ischemic stroke.[25]	Gene expression datasets	External datasets confirmed	Online datasets

Abbreviations: CADD: Computer-aided drug design; DL: Deep learning; GLMs: Generalized linear models; LASSO: Least absolute shrinkage and selection operator; LRGs: Lactate-related genes; NN: Neural network; RF: Random forest; RNA-seq: RNA sequencing; SVM: Support vector machine.

make it particularly challenging to systematically identify multi-target candidate molecules from expansive chemical libraries. AI, particularly DL and generative models, is significantly accelerating the drug discovery process for the complex pathological network of ischemic stroke through capabilities in precise prediction, intelligent *de novo* generation, and multi-objective optimization.[26,27]

3.2.1. Intelligent virtual screening and molecular docking

Virtual screening and molecular docking are core technologies for the initial screening of candidate drugs. Traditional physics-based docking methods face a fundamental compromise between accuracy and efficiency due to simplifications in their search algorithms and scoring functions.[28,29] In this context, machine learning has emerged as a powerful tool for predicting protein-ligand binding affinity. These ML models span from traditional regressors to advanced deep learning frameworks, including GNNs and transformers.[30,31] They typically incorporate multimodal inputs such as protein sequences, 3D structures, and molecular properties to enhance prediction accuracy. Specifically, artificial intelligence-based scoring functions can be trained to predict various endpoints tailored to virtual screening objectives, including binding affinity, pose quality, enrichment capacity, or to emulate computationally expensive docking scores.[32] Concurrently, rigorous evaluation of these ML-based scoring functions requires consideration of benchmark settings such as random splits and cross-target validation to avoid overestimation of performance.[29,33] For example, a recent ischemic stroke-on-a-chip model integrated with ML algorithms has been developed, enabling automated, high-throughput functional screening of drug candidates within a microphysiological system.[34] This represents an emerging trend of deep convergence between experimental validation and intelligent computation. These advanced predictive models allow for the rapid identification of target compounds from virtual libraries containing millions of molecules, concentrating resources on the most promising candidates and overcoming the scale and cost limitations inherent in traditional wet-lab methods. However, MLSFs trained on general protein–ligand datasets (*e.g.*, PDBbind, DUD-E) are not specifically designed for stroke-related targets. When applied to targets whose protein sequences share low similarity with those in the training set, their prediction accuracy decreases markedly. Therefore, when applying such models to stroke-relevant targets, careful consideration must be given to the similarity between the training set and the target of interest.[35]

In recent years, several successful cases of AI-based virtual screening in stroke drug discovery have progressed from computational predictions to experimental validation. For example, Meng *et al.*[36] developed a multi-dimensional data-driven computational drug repurposing strategy that integrated ischemic stroke transcriptomic data and neuroprotection data. By constructing a neuroprotective activity prediction model using ML algorithms, they successfully screened 19 potential neuroprotective agents from a database, ultimately focusing on sulbutiamine for comprehensive *in vitro* and *in vivo* validation. This study represents a complete closed loop from AI-driven virtual screening to multi-tiered biological validation, offering a reproducible methodological paradigm for the discovery of neuroprotective agents for ischemic stroke. Similarly, Zhou *et al.*[37] established an innovative drug discovery pipeline integrating DL. Using DL models, they predicted the bioactivities of natural compounds against seven key stroke-related targets. This pipeline successfully identified several highly active compounds, among which mulberrin and ellagic acid exhibited significant multi-target neuroprotective effects in an *in vitro* oxygen-glucose deprivation model.[37] This work demonstrates the unique advantages of deep learning in identifying multi-target natural products. However, although mulberrin and ellagic acid have shown neuroprotective effects validated in *in vitro* cell models, their *in vivo* efficacy and safety still require further investigation. Despite the significant increase in virtual screening throughput enabled by AI, its practical application still faces multiple risks of failure. First, data leakage constitutes a primary source of bias in current AI-assisted docking models.[38] When leakage occurs, it leads to overestimation of the model's true predictive utility, critically undermining both scientific validity and clinical safety.[39] Second, the overfitting of scoring functions is a widespread issue. Many deep learning-based scoring models achieve excellent performance on public benchmark datasets, yet their practical screening enrichment factors often fall short of expectations.[40,41] These modes suggest that the reliability of AI-based virtual screening depends not only on model architecture but also, more critically, on the biological similarity between the training data and the target application scenario.

3.2.2. Generative artificial intelligence and drug design

Generative AI overcomes the inherent limitations of traditional drug design methods, which rely heavily on known molecular structures, by constructing data-driven molecular generation models. These models can *de novo* design novel compounds that meet target-binding requirements and exhibit favorable drug-like properties, thereby transcending the constraints of existing chemical space.[42] Representative generative models, such as generative adversarial network (GAN), variational autoencoder (VAE), and the emerging diffusion model, learn the deep distribution patterns of known drug-like chemical space. They can autonomously generate entirely new molecular structures with desirable physicochemical properties and predicted binding activities.[43] For instance, Hu *et al.*[44] developed a model that integrates multi-view graph neural networks with a variational autoencoder, specifically designed for the efficient generation of diverse drug molecular scaffolds, enabling effective scaffold hopping and providing a novel strategy for lead optimization. Although this model was trained on a general medicinal chemistry dataset (*e.g.*, ChEMBL) and belongs to a general-purpose generative model, the aforementioned general framework is directly transferable and applicable to stroke-related compounds. However, the successful application of these models depends critically on rigorous evaluation and feedback mechanisms. Figure 3 systematically delineates the workflow of generative models in drug design for ischemic stroke, encompassing a closed-loop path from molecular generation and multidimensional metric evaluation to synthetic accessibility optimization, while also highlighting common failure modes such as mode collapse.

However, the successful application of these generative models in practice hinges upon the rigorous, multi-dimensional evaluation of their outputs and the effective mitigation of their inherent failure modes. The assessment of generative models encompasses a series of metrics, including validity (the proportion of generated molecules that are chemically valid), uniqueness (the proportion of generated molecules that are distinct from each other), novelty (the proportion of generated molecules not present in the training set), and diversity (the degree of structural variation within the generated set).[45] Optimizing solely for a single property, such as predicted binding affinity, can be misleading. This necessitates multi-objective optimization, which requires balancing potency, target selectivity, favorable ADMET profiles, and synthetic accessibility.[46] For instance, studies integrating deep learning with multi-objective particle swarm optimization to comprehensively optimize ADMET properties while maintaining bioactivity represent a valuable exploration in this direction.[47]

Despite the immense potential of generative models, the limited synthetic accessibility of many generated molecules restricts their practical application. To address this challenge, researchers have developed models that integrate modular reaction knowledge bases with reinforcement learning strategies, ensuring that the

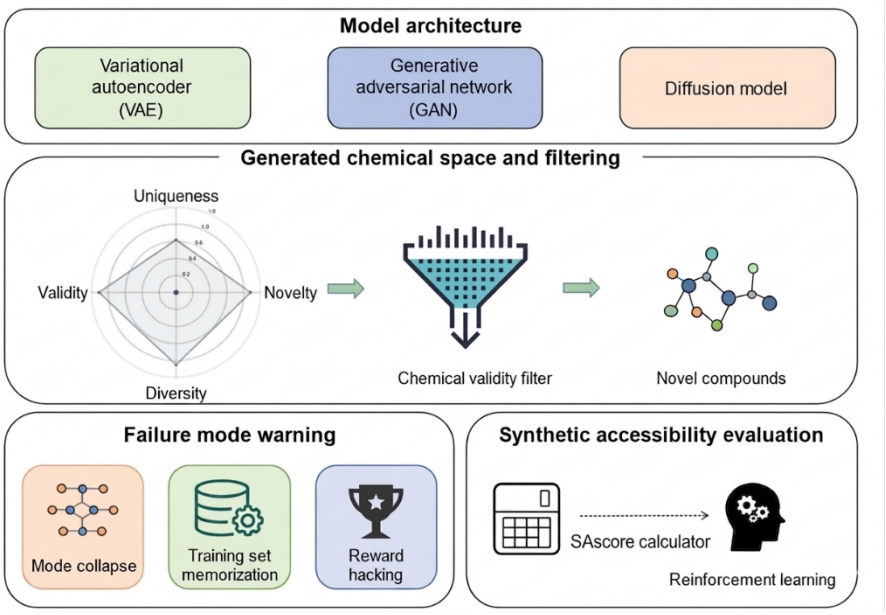


Figure 3. Workflow and multidimensional evaluation framework of generative AI for drug molecule design

proposed molecules exhibit high diversity, novelty, and strong binding propensity. In such frameworks, a reinforcement learning agent is responsible for planning feasible synthetic routes and evaluating molecular quality, thereby guiding the entire generation process. For example, the ClickGen model adopted this strategy, successfully generating, synthesizing, and validating multiple novel molecules with biological activity within just 20 days.[48] This fully demonstrates the significant potential of reinforcement learning-integrated generative design in accelerating the drug discovery cycle and enhancing research and development efficiency. This validates the ability of generative AI to significantly shorten early-stage drug discovery timelines. However, this model still has notable limitations. First, its bioactivity validation was primarily performed in tumor cell lines, and its efficacy against ischemic stroke targets has not yet been verified in pathology models relevant to ischemic stroke. Second, the lead compounds generated by the ClickGen framework have not undergone systematic pharmacokinetic (ADMET) evaluation or *in vivo* pharmacodynamic validation, leaving a considerable gap before they can be considered clinical drug candidates. Table 4 provides a consolidated overview of AI models applied to virtual screening and generative drug design in the context of ischemic stroke.

Finally, it is crucial to recognize the common failure modes of generative models. These include mode collapse (where the model generates only a limited subset of possible outputs), training set memorization (where the model reproduces molecules from its training data rather than creating novel ones), and reward hacking in reinforcement learning (where an agent learns to exploit the scoring function to achieve high scores without genuinely creating better molecules.[49,50] Only by integrating considerations of multi-objective optimization, synthetic accessibility, and these potential pitfalls can we better generate molecules that are both aligned with therapeutic goals and offer value beyond traditional approaches.

3.3. Early prediction of pharmacokinetics and toxicity (ADMET)

In ischemic stroke drug discovery, the pharmacokinetic (absorption, distribution, metabolism, and excretion) and toxicity (T) properties of candidate compounds are pivotal factors determining clinical success. The simultaneous optimization of these interconnected variables represents a major bottleneck in drug design.[51] Traditional experimental assessment heavily relies on *in vivo* animal studies, which are not only time-consuming, labor-intensive, and costly but also face significant challenges in scale and efficiency when applied to the early-stage screening of large compound libraries. Although general-purpose computational tools such as QikProp, DataWarrior, and MetaTox are already available for predicting ADMET properties, the growing demand for more innovative therapies necessitates greater accuracy, throughput, and speed in the predictive process.[52] The integration of AI, through the construction of high-performance ML and DL models, now enables the

Table 4. AI in generative artificial intelligence and drug design for ischemic stroke

No.	AI model	Application in stroke CADD	Key findings	Input data type	Validation strategy	Code/Data availability
1	MDCDR (RF/SVM/GBDT)	Drug repurposing and screening of neuroprotective agents for ischemic stroke.	Nineteen potential neuroprotective agents were screened out.[36]	Gene expression datasets	Experimental validation (*in vivo* and *in vitro*)	https://www.ncbi.nlm.nih.gov/geo/
2	SELFormer (Transformer)	Virtual screening of multi-target natural compounds and discovery of lead compounds against ischemic stroke.	Several highly active compounds were successfully identified.[37]	Small molecules in canonical SMILES format.	Molecular docking verification and in vitro validation	https://pubs.acs.org/doi/10.1021/acs.jcim.5c00135
3	ScaffoldGVAE (GNNs)	Scaffold generation and scaffold hopping for drug molecules.	ScaffoldGVAE can explore the unseen chemical space and generate novel molecules distinct from known compounds.[44]	Small molecules in canonical SMILES format.	Based on seven general evaluation metrics for generative models and four evaluation metrics specific to scaffold-hopping generative models.	https://github.com/ecust-hc/ScaffoldGVAE
4	ClickGen (DL model)	ClickGen is capable of generating highly diverse, novel, and synthetically accessible molecules.	ClickGen and related models may represent a novel paradigm for molecular generation.[48]	Compounds	Wet-lab validation	https://github.com/mywang1994/cligen_gen

Abbreviations: CADD: Computer-aided drug design; DL: Deep learning; GBDT: Gradient boosting decision tree; GNN: Graph neural network; RF: Random forest; SMILES: Simplified molecular-input line-entry system; SVM: Support vector machine.

early, rapid, and high-accuracy prediction of compound pharmacokinetic and toxicity profiles.[53]

Current AI-integrated ADMET prediction platforms can be broadly categorized into two methodological approaches: those based on ML and those based on graph computation.[54] ML-based methods utilize chemical descriptors and algorithms, such as random forest (RF) and support vector machine (SVM), to predict ADMET properties.[55-57] Their capability to identify complex patterns within large datasets has been well-validated, and they have demonstrated good applicability even for novel, complex molecular entities like targeted protein degraders. Graph-based computational methods, particularly graph neural networks (GNNs), represent the cutting edge in this field. Their core advantage lies in the ability to directly encode a molecule's topological structure, with atoms as nodes and bonds as edges, into a graph representation. Through message-passing mechanisms that aggregate neighborhood information, GNNs can fundamentally capture the local chemical environments and long-range dependencies within a molecule that determine its properties. This often enables GNNs to outperform traditional methods in predicting properties closely related to complex three-dimensional structures. For instance, research has employed GNNs to build comprehensive prediction platforms covering up to 73 ADME/T endpoints. Such platforms not only achieve high predictive accuracy but can also provide uncertainty quantification for their predictions, thereby offering more reliable guidance for the subsequent rational design and structural optimization of molecules.[58]

Furthermore, for drug development targeting neurological diseases like ischemic stroke, the ability to penetrate the blood–brain barrier is a critical parameter in the distribution phase. While general ADMET properties (*e.g.*, metabolic stability, hepatotoxicity, cardiotoxicity) are broadly applicable to any therapeutic agent, CNS-directed drug discovery necessitates particular emphasis on parameters governing blood–brain barrier penetration.[59] These include passive permeability, efflux transporter interactions (especially P-glycoprotein), unbound brain fraction (fu,brain), and neurotoxicity liabilities.[60] ML models for predicting blood–brain barrier permeability have advanced substantially, employing diverse molecular

representations and algorithms ranging from RF to deep neural networks. To address this specific challenge, ongoing research is exploring the use of large language models based on Transformer architectures to develop next-generation predictive tools. Table 5 summarizes the aforementioned models, which, through pre-training on vast corpora of compound sequences or graph structures, are capable of learning more generalized molecular representations. This capability allows for the efficient and accurate high-throughput screening of ADMET properties across extensive compound libraries, including natural products.[61] However, data scarcity remains a fundamental challenge for CNS-specific endpoints. Publicly available blood-brain barrier datasets are often limited in size and chemical diversity, and substantial variability across training datasets precludes direct comparison of models.[62] Furthermore, permeability measurements derived from different experimental assays may exhibit significant variability; meta-analyses have revealed that even within the same method, repeatability is limited, and different methods may probe distinct permeability mechanisms (*e.g.*, intrinsic versus apparent permeability).[63] Beyond endpoint-specific challenges, the reliability of ADMET predictions critically depends on uncertainty quantification and domain of applicability assessment. Recent studies have integrated conformal prediction frameworks to provide prediction intervals alongside point estimates, enabling researchers to distinguish between high-confidence predictions within the training distribution and uncertain predictions for out-of-domain compounds.[64] These uncertainty-aware approaches are essential for responsible decision-making in ischemic stroke drug discovery, where false positives in blood–brain barrier permeability prediction or false negatives in toxicity screening can lead to costly downstream failures.

3.4. Clinical trial design and patient stratification

Within the drug discovery and development pipeline, clinical trials represent a critical juncture for translating laboratory findings into clinical benefits, yet they also constitute one of the stages with the highest attrition rates.[65] Traditional stroke clinical trials frequently encounter significant challenges, including substantial patient heterogeneity, reliance on single or limited efficacy evaluation metrics, and the difficulty in accurately identifying treatment-responsive subpopulations.[66] These factors contribute to the failure of numerous therapeutic agents that demonstrated efficacy in preclinical models during subsequent clinical testing. AI is now extensively applied in clinical trials, transforming their design and execution by comprehensively enhancing trial efficiency, precision, and predictive power.[67]

The core application of AI in clinical trial design lies in enabling trial enrichment and precision patient stratification. By mining multimodal data, including neuroimaging (CT, MRI), genomics, proteomics, and electronic health records, AI models can identify pathophysiological subtypes with greater accuracy than conventional clinical classifications. This enables the enrollment of patient cohorts that are more homogeneous in terms of underlying disease mechanisms, thereby playing a crucial role in advancing precision medicine and informing clinical decision-making for ischemic stroke.[66] For instance, ML algorithms can parse diffusion-weighted imaging to automatically quantify infarct core and penumbra characteristics. This enables the selection of patients based on specific, quantifiable tissue fate profiles for trials of neuroprotective agents.[68,69] Beyond imaging, models such as IMPROVE integrate intraplaque hemorrhage features from MRI with clinical parameters

Table 5. AI in ADMET for ischemic stroke

No.	AI model	Application in stroke CADD	Key findings	Input data type	Validation strategy	Code/Data availability
1	Deep-PK (GNNs)	Evaluating pharmacokinetic properties.	Yield the best predictive performance across 73 endpoints, including 64 ADMET and 9 general properties.[58]	Small molecule datasets on endpoints	Cross-validate	https://biosig.lab.uq.edu.au/deeppk/
2	LLM	Investigate the pharmacokinetic ADME properties and blood–brain barrier permeability coefficients of compounds.	Prediction models have effectively identified the potential ADME characteristics of various compounds.[61]	Compounds	Experimental validation (*in vitro*)	https://www.biomedicinej.com/cgi/editor.cgi?article=1474&window=additional_files&context=biomedicine

Abbreviations: ADME: Absorption, distribution, metabolism, and excretion; ADMET: Absorption, distribution, metabolism, excretion, and toxicity; CADD: Computer-aided drug design; Deep-PK: Deep learning–based pharmacokinetics; GNN: Graph neural network; LLM: Large language model.

to identify individuals with symptomatic carotid artery stenosis who are at the highest risk of recurrence.[70] This provides a mechanistic rationale for the precise enrollment of patients into clinical trials for novel pharmacological agents targeting plaque stabilization.

Furthermore, AI is transforming the adjudication of endpoints and the development of digital biomarkers. DL models enable the automated and continuous analysis of serial brain imaging, accurately quantifying dynamic changes in infarct volume, hemorrhagic transformation, or subtle alterations induced by therapeutic interventions.[71,72] This process yields objective, sensitive, and reproducible early efficacy readouts. Such data-driven, AI-derived metrics are instrumental in supplementing traditional clinical scales, which often rely on subjective judgment or manual scoring. Moving a step further, AI can integrate multimodal data streams to construct novel composite endpoints.[73] These next-generation endpoints hold the potential to capture therapeutic benefits more comprehensively than single traditional endpoints and to provide greater statistical power in clinical trials.

Furthermore, predictive models can forecast patient recruitment rates and optimize trial site selection.[74] By leveraging AI to analyze interim data in real-time, recommendations can be generated for adjusting randomization ratios or for identifying emerging responder subgroups, thereby guiding adaptive trial pathways.[75,76] Evidence suggests that candidate drugs discovered or optimized with AI involvement may demonstrate higher success rates in early-phase trials, highlighting the potential of integrated AI approaches to de-risk the clinical translation pipeline.[77] Representative AI models applied in clinical trial design and patient stratification for ischemic stroke are summarized in Table 6.

4. Current status, challenges, and future perspectives

In summary, AI-assisted drug design has progressed from a theoretical concept to practical validation, permeating the entire pipeline of ischemic stroke drug discovery, from target identification to clinical trial design.[78] AI-centric platforms, leveraging ML and generative models to accelerate tasks, have significantly shortened early-stage research timelines and reduced costs.[79] However, alongside these notable advancements, translating AI-driven predictions into clinically validated effective therapies remains fraught with multiple challenges.[80] These bottlenecks hinder the full realization of AI's potential in ischemic stroke drug research and development.

The primary and fundamental obstacle lies in the data bottleneck. The performance of AI models is heavily dependent on training with high-quality, large-scale, and standardized datasets. However, in the field of stroke research, high-quality multi-omics data, standardized clinical imaging data, and comprehensive drug-target interaction data remain relatively scarce, fragmented, and heterogeneous.[81] This phenomenon limits the sufficiency and robustness of model training, potentially leading to overfitting or insufficient generalizability. To address this issue, future research should focus on constructing high-quality, specialized disease datasets for stroke to provide a solid foundation for model development.[82] Current technical approaches have adopted privacy-preserving computational techniques, such as federated learning, to enable collaborative modeling across multiple centers and institutions. This strategy effectively breaks down data silos while rigorously protecting patient privacy.[83,84] Furthermore, research employs data generation and

Table 6. AI in clinical trial design and patient stratification for ischemic stroke

No.	AI model	Application in stroke CADD	Key findings	Input data type	Validation strategy	Code/Data availability
1	EfficientNet3D	Classify stroke size in AF-related stroke.	Developed a DL algorithm that automatically classifies stroke sizes based on the imaging criteria.[68]	Diffusion-weighted imaging	Five-fold cross-validation	Not mentioned
2	IMPROVE	Identify patients with high stroke risk.	Provided accurate individual stroke risk estimates.[70]	MRI	Bootstrapping	Not mentioned

Abbreviations: AF: Atrial fibrillation; CADD: Computer-aided drug design; DL: Deep learning; MRI: Magnetic resonance imaging.

augmentation techniques, for instance, utilizing generative models to produce high-fidelity synthetic data under the premise of privacy preservation.[85,86]

Furthermore, the challenges of model interpretability and generalizability persist. The internal decision-making processes of complex AI models, particularly those based on DL, often lack clear biological or chemical mechanistic explanations.[87] In the field of drug discovery, which heavily relies on mechanistic insight, this deficiency in interpretability not only undermines experimental scientists' trust in the predictive outcomes but also poses substantial difficulties for subsequent rational molecular optimization, in-depth investigation of mechanisms of action, and compliance with the stringent requirements of regulatory agencies for drug approval.[88] To address this challenge, current research is dedicated to advancing explainable AI and exploring its integration with causal inference methods. Meanwhile, interpretable methods tailored to molecular and graph-based representations have emerged as essential bridges connecting model predictions with chemical intuition. Atom-level and bond-level attribution techniques, such as GNNs explainers or attention-based frameworks, can effectively identify substructures or functional groups that critically influence predicted properties such as binding affinity or toxicity.[89] Counterfactual explanations offer intuitive insights by generating minimal structural modifications that invert model predictions (*e.g.*, from active to inactive), thereby revealing how specific structural alterations can reverse a model's decision and providing direct guidance for lead optimization.[90] In molecular transformers such as MoLFormer, attention visualization further uncovers how models focus on spatially or chemically relevant regions of a molecule, effectively capturing long-range dependencies and atomic interactions consistent with established chemical principles.[91] Collectively, these chemically aware interpretability techniques not only enhance model transparency but also facilitate the translation of computational predictions into testable hypotheses grounded in mechanistic reasoning, thereby supporting more rational and efficient drug design in the context of ischemic stroke. This approach aims to uncover potential causal relationships between variables, rather than mere correlations, thereby providing more theoretically guided mechanistic hypotheses for drug design. For instance, in studies predicting malignant cerebral edema after ischemic stroke, researchers have utilized the SHAP method to parse the models, clearly revealing the contribution of key features, such as ischemic core volume, on the predictions.[69] Similarly, studies on hemorrhagic transformation prediction models have highlighted their clinical applicability alongside an evidence-based interpretability approach. Additionally, visualization techniques like gradient-weighted class activation mapping can map the model's attention regions when analyzing cerebrovascular images to specific anatomical structures. This bridges model decisions with clinically comprehensible morphological features, facilitating the establishment of interdisciplinary research consensus.[72,88]

Furthermore, even when model predictions perform well in computational simulations and simplified experimental settings, their translation into clinical practice remains challenging. This translational gap primarily stems from the inherent limitations of current preclinical models, including conventional cell cultures and animal models, in replicating the complex pathophysiology of human stroke.[88,92] This discrepancy directly contributes to the high attrition rate of many candidate drugs that show promise in preclinical studies but ultimately fail to demonstrate efficacy in human clinical trials. The translation from computational predictions to clinically approved therapies necessitates a systematic, multi-stage evaluation of AI-derived candidates. This process typically begins with *in silico* prediction, wherein AI models generate target hypotheses or lead compounds. It is followed by *in vitro* functional validation using high-throughput screening or organ-on-a-chip systems. Subsequent stages encompass *in vivo* models to assess efficacy and safety within complex biological systems, human biomarker validation to confirm target engagement and pharmacodynamic effects, and ultimately early-phase clinical trials to establish preliminary safety and efficacy in patients.[9,52] However, AI can also introduce pitfalls in cases where models encounter domain shift, meaning that the distribution of training data diverges from the target application context.[38] Additional risks arise in scenarios where unmeasured confounders in observational data propagate through model predictions, thereby generating spurious associations that fail to replicate in prospective settings. To mitigate this risk, the predictive validity of AI models requires iterative validation and optimization within more physiologically relevant and complex *in vitro* systems. Examples include organ-on-a-chip platforms that can mimic the blood–brain barrier or the neurovascular unit, and brain organoids that model specific regional functions.[34,93]

Beyond technical considerations, the integration and application of AI tools in clinical settings are subject to stringent regulatory and ethical frameworks. In January 2025, the U.S. Food and Drug Administration (FDA) released a draft guidance titled Considerations for the Use of Artificial Intelligence to Support Regulatory Decision-Making for Drug and Biological Products.[94]

In the same year, the FDA and the European Medicines Agency (EMA) jointly published ten guiding principles for good AI practices in drug development, emphasizing human-centered design, risk-based approaches and lifecycle-spanning governance requirements.[95] Together, these international regulatory frameworks signify that AI-assisted drug discovery is transitioning from technical exploration to standardized development, thereby providing a practical compliance pathway for AI-driven drug discovery in ischemic stroke.

At the level of data governance, privacy-preserving technologies such as federated learning have emerged as essential tools for breaking down data silos, enabling multi-institutional collaborative modeling without sharing raw data.[96] Reproducibility and auditability require that AI systems maintain clear data provenance and an immutable audit trail.[96] Furthermore, when models are applied to electronic health records (EHR) or imaging data across diverse populations, systematic efforts are needed to address bias and fairness concerns. Studies have shown that AI may underperform in underrepresented populations, necessitating continuous monitoring of demographic bias and the performance of subgroup analyses.[97] In addition, standardized documentation such as model cards and data sheets should be adopted to clearly communicate model specifications, intended use, validation performance, and limitations, thereby bridging the gap between developers and clinical end-users and supporting regulatory review.

Concurrently, the translation of AI tools from the laboratory to preclinical and clinical workflows faces multi-level practical barriers.[98] On the technical front, a significant gap exists between current capabilities and the demands of AI systems. When confronted with inconsistent input data, contextual decision-making, or nuanced regulatory requirements, AI tools often exhibit sharp performance declines.[99] Importantly, the candidate compounds that truly require breakthrough development in drug discovery are frequently situated outside existing training distributions, representing a fundamental dilemma for practical AI deployment.[100] Additionally, the integration of legacy systems, workflows, and personnel in clinical research environments severely delays data standardization, which is a prerequisite for large-scale deployment of AI models.[101,102]

In the context of stroke drug discovery, the role of AI is not to replace human experts but to serve as a powerful assistive tool.[101] A study based on over 2,000 clinical case diagnoses demonstrated that a hybrid collective system combining physicians and large language models achieved superior diagnostic accuracy compared to physician-alone, AI-alone, or simple aggregated groups, attributable to the complementarity of error types between humans and AI.[103] In the specific context of stroke drug design, AI models can efficiently screen vast chemical spaces, predict ADMET properties, and uncover cryptic structure-activity relationships, whereas human researchers are responsible for posing biologically sound scientific questions, interpreting the mechanistic implications of AI outputs, designing validation experiments, and providing theoretical explanations for unexpected results. Subsequent research should further consider how to balance the advantages of AI with these ethical and technical hurdles, which is essential for ensuring its reliability and acceptance in the context of ischemic stroke drug design.[104]

5. Conclusion

The development of therapeutic agents for ischemic stroke has long been confronted with the formidable challenges of a high attrition rate and difficulties in clinical translation. AI technologies have now permeated the entire drug discovery pipeline, ranging from intelligent target mining based on multi-omics data and structure-based virtual screening with generative molecular design, to the early prediction of pharmacokinetic properties and the optimization of clinical trial protocols. These advances have significantly accelerated the process of lead compound discovery and enhanced the efficiency of resource utilization.

However, the vast majority of AI applications are still in the early exploratory stage, and their actual clinical value awaits validation through prospective studies. To translate the computational potential of AI into clinically validated therapies, multiple obstacles must be overcome. Future progress depends on several key breakthroughs. First, it is essential to construct high-quality, standardized, and shareable multimodal stroke datasets, thereby fostering a well-functioning data ecosystem. Second, explainable AI and hybrid models that integrate physical information should be developed to enhance the reliability and biological interpretability of predictive outcomes. Concurrently, rigorous prospective validation and benchmarking studies are required to objectively assess the predictive performance of AI models. Furthermore, iterative validation and optimization should be carried out in more physiologically relevant complex *in vitro* systems, such as organ-on-a-chip platforms. Finally, there is an urgent need to establish regulatory science guidelines and ethical review principles tailored to AI-driven drug discovery.

In summary, although AI has shown promising preliminary achievements in the field of drug design for ischemic stroke, its true clinical value remains insufficiently substantiated by empirical evidence. Only by steadily

advancing along the above-mentioned directions can AI ultimately deliver truly effective therapeutic agents for patients with ischemic stroke.

Acknowledgments

None.

Funding

This work was supported by the National Natural Science Foundation of China (No. 82505804); the Hangzhou Agricultural and Social Development Key Project (No. 20231203A12); and the "Pioneer" and "Leading Goose" + "X" R&D Program of Zhejiang (No. 2025C02201).

Conflict of interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Author contributions

Conceptualization: Shutong Pang
Visualization: Yuehui Liao, Yu Hu
Writing–original draft: Shutong Pang, Yu Chen
Writing–review & editing: Shutong Pang, Xiaobo Lai, Panfei Li

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Availability of data

Not applicable.

References

1. Tao T, Liu M, Chen M, *et al*. Natural medicine in neuroprotection for ischemic stroke: Challenges and prospective. *Pharmacol Ther*. 2020;216:107695.

   doi: 10.1016/j.pharmthera.2020.107695

2. Hilkens NA, Casolla B, Leung TW, de Leeuw FE. Stroke. *Lancet*. 2024;403(10446):2820-2836.

   doi: 10.1016/s0140-6736(24)00642-1

3. Tang X, Jiang L, Luo Y, *et al*. Leukoaraiosis and acute ischemic stroke. *Eur J Neurosci*. 2021;54(6):6202-6213.

   doi: 10.1111/ejn.15406

4. Hurford R, Sekhar A, Hughes TAT, Muir KW. Diagnosis and management of acute ischaemic stroke. *Pract Neurol*. 2020;20(4):304-316.

   doi: 10.1136/practneurol-2020-002557

5. Zhao M, Wang J, Liu G, *et al*. Multi-Target and Multi-Phase Adjunctive Cerebral Protection for Acute Ischemic Stroke in the Reperfusion Era. *Biomolecules*. 20 2024;14(9) doi:10.3390/biom14091181

6. Bansal S, Sangha KS, Khatri P. Drug treatment of acute ischemic stroke. *Am J Cardiovasc Drugs*. 2013;13(1):57-69.

   doi: 10.1007/s40256-013-0007-6

7. Chamorro Á, Dirnagl U, Urra X, Planas AM. Neuroprotection in acute stroke: targeting excitotoxicity, oxidative and nitrosative stress, and inflammation. *Lancet Neurol*. 2016;15(8):869-881.

   doi: 10.1016/s1474-4422(16)00114-9

8. Li Y, Yang J, Zhang Y, Meng Q, Bender A, Chen X. Computational drug repositioning for ischemic stroke: neuroprotective drug discovery. *Future Med Chem*. 2021;13(15):1271-1283.

   doi: 10.4155/fmc-2021-0022

9. Zhong F, Xing J, Li X, *et al*. Artificial intelligence in drug design. *Sci China Life Sci*. 2018;61(10):1191-1204.

   doi: 10.1007/s11427-018-9342-2

10. Meng XF, Zheng Z, Yang L, Yang C, Li XL, Hao YF. Dynamic mechanisms and targeted interventions in cerebral ischemia-reperfusion injury: pathological cascade from ischemia to reperfusion and promising therapeutic strategies. *Front Neurosci* 2025; 18:191649533.

    doi: 10.3389/fnins.2025.1649533

11. Seo S, Lee JW. Applications of Big Data and AI-Driven Technologies in CADD (Computer-Aided Drug Design). *Methods Mol Biol*. 2024;2714:295-305.

    doi: 10.1007/978-1-0716-3441-7_16

12. Wu Z, Chen S, Wang Y, *et al*. Current perspectives and trend of computer-aided drug design: a review and bibliometric analysis. *Int J Surg*. 2024;110(6):3848-3878.

    doi: 10.1097/js9.0000000000001289

13. Zhao LL, Ciallella HL, Aleksunes LM, Zhu H. Advancing computer-aided drug discovery (CADD) by big data and data-driven machine learning modeling. *Drug Discov Today*. 2020;25(9):1624-1638.

    doi: 10.1016/j.drudis.2020.07.005

14. Khoei TT, Slimane HO, Kaabouch N. Deep learning: systematic review, models, challenges, and research directions. *Neural Comput Appl*. 2023;35(31):23103-23124.

    doi: 10.1007/s00521-023-08957-4

15. Zhang Y, Luo M, Wu P, Wu S, Lee TY, Bai C. Application of Computational Biology and Artificial Intelligence in Drug Design. *Int J Mol Sci*. 2022;23(21).

    doi: 10.3390/ijms232113568

16. Sortino M, Trovato B, Chiaramonte R, *et al*. The Role of Artificial Intelligence and Professional Expertise in Adapted Physical Activity Prescription for Orthopedic Rehabilitation. *J Funct Morphol Kinesiol.* 2026;11(1):113.

doi: 10.3390/jfmk11010113

17. Peng JJ, Zhang YY, Li RF, *et al*. Hybrid approach for drug-target interaction predictions in ischemic stroke models. *Artif Intell Med*. 2025;161:103067.

doi: 10.1016/j.artmed.2025.103067

18. Li Z, Zhang X, Ding L, *et al*. Rationale and design of the GOLDEN BRIDGE II: a cluster-randomised multifaceted intervention trial of an artificial intelligence-based cerebrovascular disease clinical decision support system to improve stroke outcomes and care quality in China. *Stroke Vasc Neurol*. 2024;9(6):723-729.

doi: 10.1136/svn-2023-002411

19. Qin C, Yang S, Chu YH, *et al*. Signaling pathways involved in ischemic stroke: molecular mechanisms and therapeutic interventions. *Signal Transduct Target Ther*. 2022;7(1):215.

doi: 10.1038/s41392-022-01064-1

20. Zhou J, Lei Y, Zhang S, Liu Y, Yi D. Panaxadiol Attenuates Neuronal Oxidative Stress and Apoptosis in Cerebral Ischemia/Reperfusion Injury via Regulation of the JAK3/STAT3/HIF-1α Signaling Pathway. *CNS Neurosci Ther*. 2025;31(2):e70233.

doi: 10.1111/cns.70233

21. Pun FW, Ozerov IV, Zhavoronkov A. AI-powered therapeutic target discovery. *Trends Pharmacol Sci*. 2023;44(9):561-572.

doi: 10.1016/j.tips.2023.06.010

22. Jia J, Niu L, Feng P, *et al*. Identification of Novel Biomarkers for Ischemic Stroke Through Integrated Bioinformatics Analysis and Machine Learning. *J Mol Neurosci*. 2025;75(1):13.

doi: 10.1007/s12031-025-02309-8

23. Hei C, Li X, Wang R, *et al*. Machine learning analysis of gene expression profiles of pyroptosis-related differentially expressed genes in ischemic stroke revealed potential targets for drug repurposing. *Sci Rep*. 2025;15(1):7035.

doi: 10.1038/s41598-024-83555-5

24. Liu L, Cai Y, Deng C. Identification of ANXA3 as a biomarker associated with pyroptosis in ischemic stroke. *Eur J Med Res*. 2023;28(1):596.

doi: 10.1186/s40001-023-01564-y

25. Mo J, Ning D, Chen S, *et al*. Construction of a diagnostic prediction model for ischemic stroke using lactylation-related genes. *Front Mol Neurosci*. 2025;18:1672374.

doi: 10.3389/fnmol.2025.1672374

26. Gupta R, Srivastava D, Sahu M, Tiwari S, Ambasta RK, Kumar P. Artificial intelligence to deep learning: machine intelligence approach for drug discovery. *Mol Divers*. 2021;25(3):1315-1360.

doi: 10.1007/s11030-021-10217-3

27. Doytchinova I. Drug design—past, present, future. *Molecules*. 2022;27(5):1496.

doi: 10.3390/molecules27051496

28. Ferreira LG, Dos Santos RN, Oliva G, Andricopulo AD. Molecular docking and structure-based drug design strategies. *Molecules*. 2015;20(7):13384-421.

doi: 10.3390/molecules200713384

29. Jastrzębski S, Szymczak M, Pocha A, *et al*. Emulating Docking Results Using a Deep Neural Network: A New Perspective for Virtual Screening. *J Chem Inf Model*. 2020;60(9):4246-4262.

doi: 10.1021/acs.jcim.9b01202

30. Gu S, Shen C, Zhang X, *et al*. Benchmarking AI-powered docking methods from the perspective of virtual screening. *Nat Mach Intell*. 2025;7(3):509-520.

doi: 10.1038/s42256-025-00993-0

31. Cao Z, Sciabola S, Wang Y. Large-Scale Pretraining Improves Sample Efficiency of Active Learning-Based Virtual Screening. *J Chem Inf Model*. 2024;64(6):1882-1891.

doi: 10.1021/acs.jcim.3c01938

32. Jiang J, Li D, Wang G, Wei GW. Recent advances in machine learning predictions of protein-ligand binding affinities. *Curr Opin Struct Biol*. 2026;96:103193.

doi: 10.1016/j.sbi.2025.103193

33. Yang L, Yang G, Chen X, *et al*. Deep Scoring Neural Network Replacing the Scoring Function Components to Improve the Performance of Structure-Based Molecular Docking. *ACS Chem Neurosci*. 2021;12(12):2133-2142.

doi: 10.1021/acschemneuro.1c00110

34. Liu J, Wang P, Lian P, Qin J. An ischemic stroke-on-a-chip model integrated with machine learning for screening of drug candidates. *Lab Chip*. 2025;25(24):6517-6532.

doi: 10.1039/d5lc00517e

35. Ashtawy HM, Mahapatra NR. Machine-learning scoring functions for identifying native poses of ligands docked to known and novel proteins. *BMC Bioinform*. 2015;16 Suppl 6(Suppl 6):S3.

doi: 10.1186/1471-2105-16-s6-s3

36. Meng Q, Liu Q, Mi Y, *et al*. Multi-dimensional data-driven computational drug repurposing strategy for screening novel neuroprotective agents in ischemic stroke. *Theranostics*. 2025;15(15):7653-7676.

doi: 10.7150/thno.112608

37. Zhou J, Li C, Yue Y, Kim YK, Park S. Multitarget Natural Compounds for Ischemic Stroke Treatment: Integration of Deep Learning Prediction and Experimental Validation. *J Chem Inf Model*. 2025;65(7):3309-3323.

doi: 10.1021/acs.jcim.5c00135

38. Kapoor S, Narayanan A. Leakage and the reproducibility crisis in machine-learning-based science. *Patterns*. 2023;4(9):100804.

doi: 10.1016/j.patter.2023.100804

39. Starcke J, Spadafora J, Spadafora J, Spadafora P, Toma M. The Effect of Data Leakage and Feature Selection on Machine Learning Performance for Early Parkinson's Disease Detection. *Bioengineering*. 2025;12(8):845.

doi: 10.3390/bioengineering12080845

40. Chen L, Blay V, Ballester PJ, Houston DR. SCORCH2: A Generalized Heterogeneous Consensus Model for High-Enrichment Interaction-Based Virtual Screening. *Adv Sci*. 2025;12(42):e08318.

doi: 10.1002/advs.202508318

41. Quiroga R, Villarreal MA. Developing Generalizable Scoring Functions for Molecular Docking: Challenges and Perspectives. *Curr Med Chem*. 2025;32(28):5960-5972.

doi: 10.2174/0109298673334469241017053508

42. Ocana A, Pandiella A, Privat C, *et al*. Integrating artificial intelligence in drug discovery and early drug development: a transformative approach. *Biomark Res*. 2025;13(1):45.

doi: 10.1186/s40364-025-00758-2

43. Jayasrilakshmi S, Mahapatra A. Generative AI in medicine: A thorough examination of applications, challenges, and future perspectives. *Comput Biol Med*. 2026;203:111469.

doi: 10.1016/j.compbiomed.2026.111469

44. Hu C, Li S, Yang C, *et al*. ScaffoldGVAE: scaffold generation and hopping of drug molecules via a variational autoencoder based on multi-view graph neural networks. *J Cheminform*. 2023;15(1):91.

doi: 10.1186/s13321-023-00766-0

45. Özçelik R, Grisoni F. How evaluation choices distort the outcome of generative drug discovery. *J Cheminform*. 2025;17(1):169.

doi: 10.1186/s13321-025-01108-y

46. van den Broek RL, Patel S, van Westen GJP, Jespers W, Sherman W. In Search of Beautiful Molecules: A Perspective on Generative Modeling for Drug Design. *J Chem Inf Model*. 2025;65(18):9383-9397.

doi: 10.1021/acs.jcim.5c01203

47. Yi JC, Yang ZY, Zhao WT, *et al*. ChemMORT: an automatic ADMET optimization platform using deep learning and multi-objective particle swarm optimization. *Brief Bioinform*. 2024;25(2)doi:10.1093/bib/bbae008

48. Wang M, Li S, Wang J, *et al*. ClickGen: Directed exploration of synthesizable chemical space via modular reactions and reinforcement learning. *Nat Commun*. 2024;15(1):10127.

doi: 10.1038/s41467-024-54456-y

49. Bhowmik D, Zhang P, Fox Z, Irle S, Gounley J. Enhancing molecular design efficiency: Uniting language models and generative networks with genetic algorithms. *Patterns*. 2024;5(4):100947.

doi: 10.1016/j.patter.2024.100947

50. Yoshizawa T, Ishida S, Sato T, Ohta M, Honma T, Terayama K. A data-driven generative strategy to avoid reward hacking in multi-objective molecular design. *Nat Commun*. 2025;16(1):2409.

doi: 10.1038/s41467-025-57582-3

51. Dulsat J, López-Nieto B, Estrada-Tejedor R, Borrell JI. Evaluation of Free Online ADMET Tools for Academic or Small Biotech Environments. *Molecules*. 2023;28(2)776.

doi: 10.3390/molecules28020776

52. Ferreira LLG, Andricopulo AD. ADMET modeling approaches in drug discovery. *Drug Discov Today*. 2019;24(5):1157-1165.

doi: 10.1016/j.drudis.2019.03.015

53. Fan NN, Chen J, Wang JH, Chen ZS, Yang YF. Bridging data and drug development: Machine learning approaches for next-generation ADMET prediction. *Drug Discov Today*. 2025;30(11)104487.

doi: 10.1016/j.drudis.2025.104487

54. Zhang JY, Li HL, Zhang YC, *et al*. Computational toxicology in drug discovery: applications of artificial intelligence in ADMET and toxicity prediction. *Brief Bioinform*. 2025;26(5) bbaf533.

doi: 10.1093/bib/bbaf533

55. Peteani G, Huynh MTD, Gerebtzoff G, Rodríguez-Pérez R. Application of machine learning models for property prediction to targeted protein degraders. *Nat Commun*. 2024;15(1):5764.

doi: 10.1038/s41467-024-49979-3

56. Fang C, Wang Y, Grater R, *et al*. Prospective Validation of Machine Learning Algorithms for Absorption, Distribution, Metabolism, and Excretion Prediction: An Industrial Perspective. *J Chem Inf Model*. 2023;63(11):3263-3274.

doi: 10.1021/acs.jcim.3c00160

57. Han H, Shaker B, Lee JH, *et al*. Employing Automated Machine Learning (AutoML) Methods to Facilitate the In Silico ADMET Properties Prediction. *J Chem Inf Model*. 2025;65(7):3215-3225.

doi: 10.1021/acs.jcim.4c02122

58. Myung Y, de Sá AGC, Ascher DB. Deep-PK: deep learning for small molecule pharmacokinetic and toxicity prediction. *Nucleic Acids Res*. 2024;52(W1):W469-w475.

doi: 10.1093/nar/gkae254

59. Pyka P, Garbo S, Murzyn A, *et al*. Unlocking the potential of higher-molecular-weight 5-HT(7)R ligands: Synthesis, affinity, and ADMET examination. *Bioorg Chem*. 2024;151:107668.

doi: 10.1016/j.bioorg.2024.107668

60. Blicharz-Futera K, Kamiński M, Grychowska K, Canale V, Zajdel P. Current development in sulfonamide derivatives to enable CNS-drug discovery. *Bioorg Chem*. 2025;156:108076.

doi: 10.1016/j.bioorg.2024.108076

61. Yang JS, Huang ET, Liao KY, *et al*. Artificial intelligence-driven prediction and validation of blood-brain barrier permeability and absorption, distribution, metabolism, excretion profiles in natural product research laboratory compounds. *Biomedicine*. 2024;14(4):82-91.

doi: 10.37796/2211-8039.1474

62. Meng F, Xi Y, Huang J, Ayers PW. A curated diverse molecular database of blood-brain barrier permeability with chemical descriptors. *Sci Data*. 2021;8(1):289.

doi: 10.1038/s41597-021-01069-5

63. Storchmannová K, Balouch M, Juračka J, Štěpánek F, Berka K. Meta-Analysis of Permeability Literature Data Shows Possibilities and Limitations of Popular Methods. *Mol Pharm*. 2025;22(3):1293-1304.

doi: 10.1021/acs.molpharmaceut.4c00975

64. Li P, Hua L, Ma Z, Hu W, Liu Y, Zhu J. Conformalized Graph Learning for Molecular ADMET Property Prediction and Reliable Uncertainty Quantification. *J Chem Inf Model*. 2024;64(23):8705-8717.

doi: 10.1021/acs.jcim.4c01139

65. Yildirim Z, Swanson K, Wu X, Zou J, Wu J. Next-Gen Therapeutics: Pioneering Drug Discovery with iPSCs, Genomics, AI, and Clinical Trials in a Dish. *Annu Rev Pharmacol Toxicol*. 2025;65(1):71-90.

doi: 10.1146/annurev-pharmtox-022724-095035

66. Ding L, Mane R, Wu Z, *et al*. Data-driven clustering approach to identify novel phenotypes using multiple biomarkers in acute ischaemic stroke: A retrospective, multicentre cohort study. *eClinicalMedicine*. 2022;53:101639.

doi: 10.1016/j.eclinm.2022.101639

67. Aliper A, Kudrin R, Polykovskiy D, *et al*. Prediction of Clinical Trials Outcomes Based on Target Choice and Clinical Trial Design with Multi-Modal Artificial Intelligence. *Clin Pharmacol Ther*. 2023;114(5):972-980.

doi: 10.1002/cpt.3008

68. Kim H, Kim JT, Lee H, *et al*. Deep Learning-Based Automatic Classification of Stroke Size in Patients With Atrial Fibrillation. *J Stroke*. 2025;27(3):405-408.

doi: 10.5853/jos.2025.00423

69. Dong Z, Huang Z, Hu H, *et al*. Development and validation of an interpretable machine learning model to predict malignant cerebral edema after endovascular treatment in acute anterior circulation large vessel occlusion stroke. *Front Neurol*. 2026;161694030.

doi: 10.3389/fneur.2025.1694030

70. Nies KPH, Smits LJM, van Kuijk SMJ, *et al*. Individualized MRI-Based Stroke Prediction Score Using Plaque Vulnerability for Symptomatic Carotid Artery Disease Patients (IMPROVE). *Stroke*. 2025;56(8):2068-2078.

doi: 10.1161/strokeaha.124.050020

71. Abraham S, Carone D, Mielke O, *et al*. Automatically quantified follow-up imaging biomarkers predict clinical outcomes after acute ischemic stroke. *Front Neurol*. 2025;16:1483138.

doi: 10.3389/fneur.2025.1483138

72. Tao X, Shen S, Yang L, *et al*. Clinically oriented deep learning framework for automated vessel wall segmentation in black-blood MRI: a multi-center study. *Eur Radiol*. 2025.

doi: 10.1007/s00330-025-12161-4

73. Santo B, Poppenberg K, Ciecierska S-S, *et al*. Machine Learning Models Incorporating Clot Histomics and Transcriptomics Predict Functional Outcome After Mechanical Thrombectomy in Acute Ischemic Stroke. Meeting Abstract. *Neurosurgery*. 2025;71:235-235.

doi: 10.1227/neu.0000000000003360_1379

74. Taylor A, Castle A, Merino JG, Hsia A, Kidwell CS, Warach S. Optimizing stroke clinical trial design: estimating the proportion of eligible patients. *Stroke*. 2010;41(10):2236-8.

doi: 10.1161/strokeaha.110.578252

75. Park Y, Nycklemoe S. MARGO: Machine Learning-Assisted Adaptive Randomization for Group Sequential Trials Based on Overlap Weights. *Stat Med*. 2025;44(15-17):e70158.

doi: 10.1002/sim.70158

76. Mangalam M. AI-driven dynamic grouping for adaptive clinical trials: Rethinking randomization in precision medicine. *Artif Intell Med*. 2025;170:103272.

doi: 10.1016/j.artmed.2025.103272

77. Kp Jayatunga M, Ayers M, Bruens L, Jayanth D, Meier C. How successful are AI-discovered drugs in clinical trials? A first analysis and emerging lessons. *Drug Discov Today*. 2024;29(6):104009.

doi: 10.1016/j.drudis.2024.104009

78. Vora N, Shah S, Patel P, Shah M. Artificial intelligence and

multi-omics in drug discovery: A deep learning-powered revolution. *Cure Care*. 2025:100011.

doi: 10.1016/j.ccwv.2025.100011

79. Dharmasivam M, Kaya B, Akinware A, Azad MG, Richardson DR. Leading artificial intelligence-driven drug discovery platforms: 2025 landscape and global outlook. *Pharmacol Rev*. 2026;78(1):100102.

doi: 10.1016/j.pharmr.2025.100102

80. Bassey GE, Daniel EA, Okesina KB, Odetayo AF. Transformative Role of Artificial Intelligence in Drug Discovery and Translational Medicine: Innovations, Challenges, and Future Prospects. *Drug Des Devel Ther*. 2025;19:7493-7502.

doi: 10.2147/dddt.S538269

81. Paul D, Sanap G, Shenoy S, Kalyane D, Kalia K, Tekade RK. Artificial intelligence in drug discovery and development. *Drug Discov Today*. 2021;26(1):80-93.

doi: 10.1016/j.drudis.2020.10.010

82. Wang W, Kiik M, Peek N, *et al*. A systematic review of machine learning models for predicting outcomes of stroke with structured data. *PLoS ONE*. 2020;15(6):e0234722.

doi: 10.1371/journal.pone.0234722

83. Razfar N, Kashef R, Mohammadi F. PSA-FL-CDM: A Novel Federated Learning-Based Consensus Model for Post-Stroke Assessment. *Sensors*. 2024;24(16)5095.

doi: 10.3390/s24165095

84. Yang J, Xie H, Huang L, Gao Z, Shen S. Predicting the Prognosis of Stroke Patients Based on Personalized Federated Learning. *J Internet Technol*. 2024;25(6):815-824.

doi: 10.70003/160792642024112506002

85. Hadley AJ, Pulliam CL. Enhancing Activity Recognition After Stroke: Generative Adversarial Networks for Kinematic Data Augmentation. *Sensors*. 2024;24(21)6861.

doi: 10.3390/s24216861

86. Wang Y, Wang K, Hou J, Ye Z, Lin C, Li X. A modified generative adversarial networks method for assisting the diagnosis of deep venous thrombosis complications in stroke patients. *Sci Rep*. 2025;15(1)22372.

doi: 10.1038/s41598-025-04880-x

87. Islam MS, Hussain I, Rahman MM, Park SJ, Hossain MA. Explainable Artificial Intelligence Model for Stroke Prediction Using EEG Signal. *Sensors*. 2022;22(24).

doi: 10.3390/s22249859

88. Alqadi MM, Vidal SGM. Artificial Intelligence in Vascular Neurology: Applications, Challenges, and a Review of AI Tools for Stroke Imaging, Clinical Decision Making, and Outcome Prediction Models. *Curr Neurol Neurosci Rep*. 2025;25(1):34.

doi: 10.1007/s11910-025-01422-w

89. Chen Y, Wang J, Zou Q, *et al*. DrugDAGT: a dual-attention graph transformer with contrastive learning improves drug-drug interaction prediction. *BMC Biol*. 2024;22(1):233.

doi: 10.1186/s12915-024-02030-9

90. Lamens A, Bajorath J. Generation of Molecular Counterfactuals for Explainable Machine Learning Based on Core-Substituent Recombination. *ChemMedChem*. 2024;19(3):e202300586.

doi: 10.1002/cmdc.202300586

91. Maziarka Ł, Majchrowski D, Danel T, *et al*. Relative molecule self-attention transformer. *J Cheminform*. 2024;16(1):3.

doi: 10.1186/s13321-023-00789-7

92. Kunapinun A, Suthakorn J, Siravaman D, Songsaeng D. Toward AI-Powered Neurovascular Intervention: From Imaging to XR-Robotic Convergence. *Stroke*. 2025.

doi: 10.1161/strokeaha.125.053121

93. Lombardozzi G, Szebényi K, Giorgi C, *et al*. Next-Gen Stroke Models: The Promise of Assembloids and Organ-on-a-Chip Systems. *Cells*. 2025;14(24):1986.

doi: 10.3390/cells14241986

94. Deshmukh AD, Wagner JK. FDA Draft Guidelines for AI and the Need for Ethical Frameworks. *JAMA Pediatr*. 2025;179(9):937-938.

doi: 10.1001/jamapediatrics.2025.1979

95. Morino E, Tokita D. AI in clinical trials: Current status, challenges, and future directions for emergency infectious disease clinical trials -Insights from the 2025 iCROWN Symposium. *Glob Health Med*. 2026;8(1):70-71.

doi: 10.35772/ghm.2026.01014

96. Singh R, Zhou K, Auclair JR. Reimagining drug regulation in the age of AI: a framework for the AI-enabled Ecosystem for Therapeutics. *Front Med*. 2025;12:1679611.

doi: 10.3389/fmed.2025.1679611

97. Yi PH, Bachina P, Bharti B, *et al*. Pitfalls and Best Practices in Evaluation of AI Algorithmic Biases in Radiology. *Radiology*. 2025;315(2):e241674.

doi: 10.1148/radiol.241674

98. Ponce-Bobadilla AV, Bräm D, Farnoud A, *et al*. Predictive AI in Clinical Pharmacology: A Call to Action to Develop Robust Benchmarking Practices. *CPT Pharmacometrics Syst Pharmacol*. 2026;15(1):e70155.

doi: 10.1002/psp4.70155

99. van Tilborg D, Rossen L, Grisoni F. Molecular deep learning at the edge of chemical space. *Nat Mach Intell*. 2026;8(4):575-587.

doi: 10.1038/s42256-026-01216-w

100. Ghislat G, Hernandez-Hernandez S, Piyawajanusorn C, Ballester PJ. Data-centric challenges with the application and adoption of artificial intelligence for drug discovery. *Expert Opin Drug Discov*. 2024;19(11):1297-1307.

doi: 10.1080/17460441.2024.2403639

101. Friebe M. AI in radiology and interventions: a structured narrative review of workflow automation, accuracy, and efficiency gains of today and what's coming. *Int J Comput Assist Radiol Surg*. 2026;21(1):1-10.

doi: 10.1007/s11548-025-03547-2

102. Furukawa S, Uchida H, Kishimoto T. Artificial Intelligence in Drug Discovery and Development: Raising Quality per Decision. *Pharmacopsychiatry*. 2026;59(03):103-116.

doi: 10.1055/a-2810-8972

103. Zöller N, Berger J, Lin I, *et al*. Human-AI collectives most accurately diagnose clinical vignettes. *Proc Natl Acad Sci USA*. 2025;122(24):e2426153122.

doi: 10.1073/pnas.2426153122

104. Daidone M, Ferrantelli S, Tuttolomondo A. Machine learning applications in stroke medicine: advancements, challenges, and future prospectives. *Neural Regen Res*. 2024;19(4):769-773.

doi: 10.4103/1673-5374.382228