A-JMRHS

AI-DRIVEN PARADIGMS IN PRECISION ONCOLOGY: MAPPING THE CONVERGENCE OF MULTI-OMICS DRUG DISCOVERY, MEDICINAL CHEMISTRY, AND ENVIRONMENTAL CARCINOGENESIS

Sanyogita Shahi[1*], Shirish Kumar Singh[2], Vinod Kumar Choudhary[3]
[1*]Department of Chemistry, Kalinga University, Raipur, Chhattisgarh 492101, India, 0000-0002-0040-1600.
[2]Regional Science Centre, Daldal Seoni, Saddu, Raipur, Chhattisgarh, 492014, India, 0009-0006-7471-5318.
[3]Department of Environmental Science, Dr R M L Awadh University, Ayodhya, 224001, India, 0000-00015675-7270.
Corresponding Author: Sanyogita Shahi[1]
Department of Chemistry, Kalinga University, Raipur, Chhattisgarh 492101, India, 0000-0002-0040-1600.
Email: drsanyogitashahi@gmail.com

ABSTRACT
Traditional oncology drug discovery is severely constrained by prolonged timelines, high attrition rates, and the non-linear, spatial-temporal heterogeneity of malignant neoplasms. Rapid mutational rewiring and the active upregulation of ATP-binding cassette (ABC) efflux transporters (e.g., ABCB1) routinely undermine static small-molecule pipelines. Artificial intelligence (AI) has emerged as a multi-scale systems biology framework capable of transforming this landscape from empirical screening into predictive *in silico* multi-parameter optimization. This systematic review comprehensively evaluates computational advancements in oncology drug discovery published between 1997 and 2026. We examine the structural parameterisation of multi-modal data layouts—including 1D SMILES, 3D molecular graphs, 3D voxel density grids, and single-cell transcriptomics—across four core analytical pillars: target discovery, *de novo* molecular generation, virtual screening, and precision clinical translation. Furthermore, we audit the field's primary computational and ethical bottlenecks through the lens of algorithmic fairness, socio-demographic training bias, data-silo constraints, and hardware-aware cryptographic privacy.
The evidence synthesizes how deep learning architectures (such as Graph Neural Networks, Generative Adversarial Networks, and Vision Transformers) accurately model biochemical cascades, bypass costly semi-empirical wavefunction simulations, and map the metabolic bioactivation pathways of exogenous environmental carcinogens. However, significant challenges remain regarding dataset shift and demographic skew, where public biobanks overrepresent European ancestry (81.3%), leading to elevated predictive errors (up to 34.5%) in underserved cohorts. We evaluate technical solutions to these constraints, documenting that fairness-aware loss optimization successfully reduces performance variances across ancestral subgroups. Additionally, decentralized architectures—specifically Federated Learning combined with Homomorphic Encryption (HE) or Secure Multi-Party Computation (MPC)—demonstrate generalizability scores matching centralized data pools while preserving patient privacy. AI has transitioned into an indispensable, data-driven framework for modern oncology chemistry and structural toxicology. Realizing its full clinical potential requires moving beyond internal cross-validation toward hybrid workflows that combine physics-informed neural networks with prospective multi-centric clinical validation and robust, localized bias-mitigation layers.

Keywords: Artificial Intelligence, Cancer Drug Discovery, Machine Learning, Deep Learning, Virtual Screening, Precision Oncology, Drug Repurposing.

INTRODUCTION

Structural, Quantum-Chemical, and Biochemical Heterogeneity in Oncological Drug Discovery

www.ajmrhs.com
eISSN: 2583-7761

Date of Received: 06-06-2026
Date Acceptance: 16-06-2026
Date of Publication: 15-07-2026

The architectural landscape of modern oncological drug discovery is constrained by the non-linear, spatial-temporal heterogeneity of malignant neoplasms [1]. At the molecular level, tumor progression is driven by a chaotic accumulation of somatic mutations, structural variations, and epigenetic remodeling that yields highly individualized cellular phenotypes [1][2]. This diversity undermines traditional, static small-molecule design pipelines [2]. Candidate therapeutics targeting specific oncogenic proteins—

such as the epidermal growth factor receptor (EGFR) or the anaplastic lymphoma kinase (ALK)—frequently experience sharp declines in clinical efficacy due to rapid mutational rewiring within the adenosine triphosphate (ATP) binding pocket or allosteric regulatory domains [2][3].

From a biochemical perspective, tumor resistance mechanisms represent a highly coordinated system designed to evade exogenous xenobiotics [3][4]. A major manifestation of this is the upregulation of ATP-binding cassette (ABC) transporter proteins, notably P-glycoprotein (P-gp/ABCB1), multidrug resistance-associated protein 1 (MRP1/ABCC1), and breast cancer resistance protein (BCRP/ABCG2) [4]. These cross-membrane efflux pumps utilize the free energy of ATP hydrolysis (△G ATP approx -30.5 kJ/mol) to actively transport structurally diverse hydrophobic chemotherapeutic agents out of the intracellular cytoplasm, as shown mathematically by Figure 1.1.

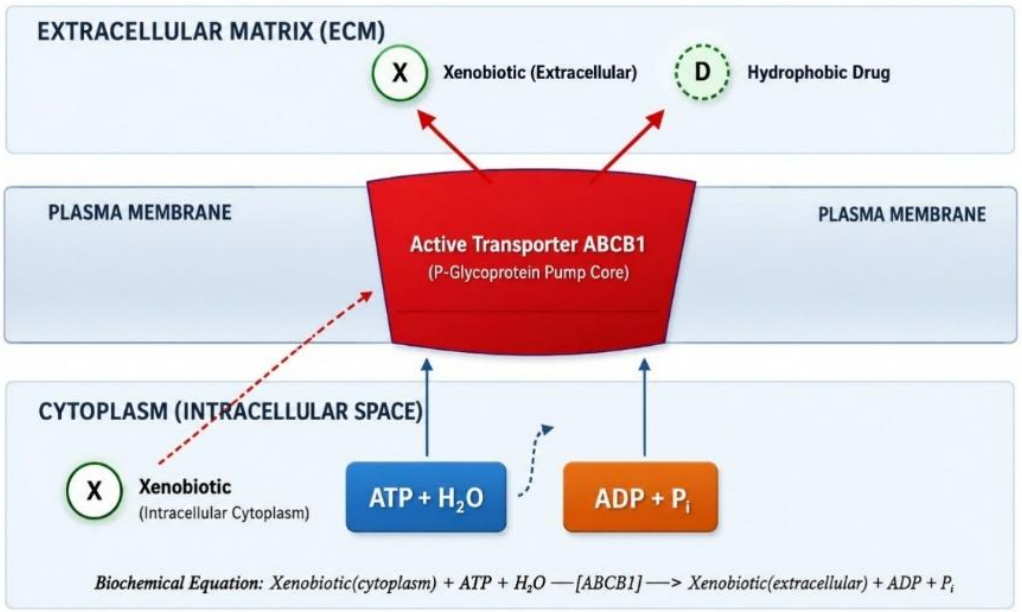


FIGURE 1.1: Mechanistic Block Diagram of ABCB1 (P-Glycoprotein) Mediated Xenobiotic Efflux

The preclinical validation pipeline for oncology assets is slowed by high attrition rates during the translation from *in vitro* high-throughput screening (HTS) to *in vivo* pharmacokinetic and pharmacodynamic (PK/PD) profiling [5][6]. Small molecules optimized strictly for thermodynamic binding affinity (such as low dissociation constants, Kd $\leq 10^{-9}$ M) often exhibit poor absorption, distribution, metabolism, excretion, and toxicity (ADMET) characteristics [6].

These computational and physiological failures typically stem from unoptimized lipophilicity profiles (log P values outside the Lipinski rule of five envelope of 0 <log P < 5), high clearance rates by hepatic cytochrome P450 (CYP) monooxygenases, or unexpected off-target toxicities resulting from structural cross-reactivity with normal somatic proteins [6][7]. Consequently, developing predictive *in silico* chemical frameworks is an essential strategy for de-risking lead scaffolds before synthetic execution and *in vitro* biological assays [7].

The Convergence of Cheminformatics, Environmental Xenobiotics, and Artificial Intelligence

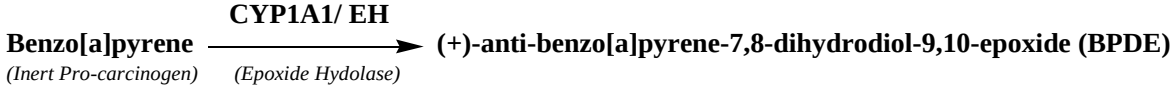
Artificial intelligence (AI) serves as a core computational architecture in precision chemistry by converting unstructured chemical and biological data into predictable, non-linear deep-learning feature maps [8]. Rather than relying on classical quantitative structure-activity relationship (QSAR) models that utilize static, linear descriptor parameters, deep neural networks (DNNs) ingest multi-modal data structures simultaneously [8][9]. These include topological Simplified Molecular Input Line Entry System (SMILES) strings, molecular graphs (G = (V, E) where vertices V map atomic nodes and edges E designate chemical bonds), 3D protein-ligand crystallographic density grids, transcriptomic expression matrices, and tissue-level digital histopathology slides [9][10]. Table 1 summarizes how these variable chemical and biological inputs are structurally parameterized [**Table 1.1**] within modern machine learning pipelines [10].

Table 1.1: Structural Parameterization of Multi-Modal Data Layouts Inside Oncology AI Engines

S. No.	Computational Modality	Data Structure Format	Primary Mathematical Target Representation	Deep Learning Architecture	Modeled Biochemical Endpoint
1.	Topological Cheminformatics	1d Smiles/ Ascii Strings	Adjacency Matrices & Local Node Atom Vectors	Graph Neural Networks (Gnn) / Transformers	Target Binding Energy (△G Bind)
2.	Quantum Chemistry	3d Voxel Coordinate Grids	Electron Density Maps & Electrostatic Potentials	3d Convolutional Neural Networks (3d-Cnn)	Quantum Dipole Moment (Mu) & Active Pocket Fitting
3.	Functional Genomics	2d Continuous Real Vectors	Transcriptomic Rna-Seq Abundance Arrays	Variational Autoencoders (Vae)/Multilayer Perceptrons	Ic50 Dose Response/ Synthetic Lethality Induction
4.	Digital Histology	Spatial Tensor Tiled Matrices	Pixel-Level Tissue Architecture Rgb Convolutions	Deep Resnet / Vision Transformers (Vit)	Phenotypic Spatial Drug Resistance Profiling

This computational framework connects core biochemistry with environmental toxicology by identifying how exogenous carcinogens interact with endogenous somatic structures [11]. Deep neural network architectures can model the metabolic bioactivation pathways of halogenated persistent organic pollutants (POPs) and polycyclic aromatic hydrocarbons (PAHs), including benzo[a]pyrene [11].

AI models accurately map how phase I biotransformation enzymes, primarily the cytochrome P450 (CYP1A1) isoform, catalyse the oxidation of inert pro-carcinogens into highly reactive, electrophilic vicinal diol epoxides [12]. Metabolic bio-activation & chemical topology of polycyclic aromatic hydrocarbons:



The resulting electrophilic intermediates readily undergo nucleophilic attack by the exocyclic amino groups of deoxyguanosine residues within DNA molecules [12][13]. This forms covalent, bulky trans-addition DNA adducts that disrupt the high-fidelity operation of DNA polymerases, inducing characteristic G to T transversion mutations within critical oncogenic driver nodes such as the tumor suppressor protein TP53 and the GTPase KRAS [13][14].

By processing these multi-layered toxicological vectors, AI platforms can guide fragment-based *de novo* molecular design, optimize virtual screening libraries, and isolate chemical biomarkers that dictate structural tissue sensitivity to targeted anticancer compounds [14].

Computational Mapping from Early Chemical Screening to Clinical Translation

The deployment of deep machine learning architectures allows for systemic optimization across the drug discovery pipeline, from initial hits to clinical validation [15]. During the virtual screening stage, generative models—such as Reinforcement Learning-driven De Novo Molecular Generation engines and Generative Adversarial Networks (GANs)—sample vast regions of chemical space (sim 10^{60} potential drug-like molecules) to locate novel structural leads [16].

These workflows evaluate structures against target binding pockets, such as the ATP-binding cleft of mutated EGFR or mutated variants of the BRAF kinase [16]. **Figure 1.2** illustrates the structural chemical configuration of Gefitinib, an established quinazoline-core EGFR tyrosine kinase inhibitor, alongside its structural line representation [17].

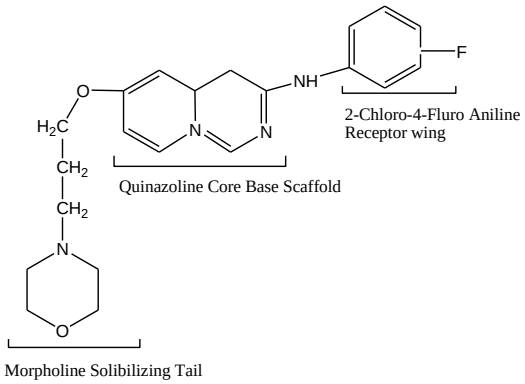


FIGURE 1.2: Molecular Pharmacophore & Structural Schematic of GEFITINIB

During the clinical translation phase, deep-learning models aid in patient stratification by matching a molecule's structural mechanism of action with the molecular multi-omics profiles of prospective clinical trial cohorts [17]. By analyzing transcriptomic signature shifts, chromatin accessibility landscapes (ATAC-Seq), and proteomic abundance matrices under drug stress, these systems can predict potential resistance mutations before patients are enrolled in clinical studies [18]. This helps reduce the rate of clinical trial failures, minimize late-stage toxicity signals, and optimize real-world survival outcomes across patient populations [19].

Scientific Rationale and Scope of Advanced Machine Learning in Chemistry

The integration of machine learning into modern oncology chemistry represents a fundamental shift toward data-driven molecular mechanics and structural toxicology [20]. By mathematically mapping quantum chemical descriptions—such as frontier molecular orbital energies (epsilon LUMO – epsilon HOMO gap), isotropic polarizabilities, and atomic Mulliken charges—directly to biological responses, AI bypasses the time-consuming process of setting up explicit, non-linear semi-empirical wavefunctions or running long, unconstrained molecular dynamics (MD) trajectories [21].

The predictive validity of these pipelines depends on data consistency, structural explainability via Physics-Informed Neural Networks (PINNs), and confirmation through wet-lab *in vitro* and *in vivo* testing [22]. **Figure 1.3** highlights the primary analytical components that constitute a modern machine learning workflow for structural discovery [22].

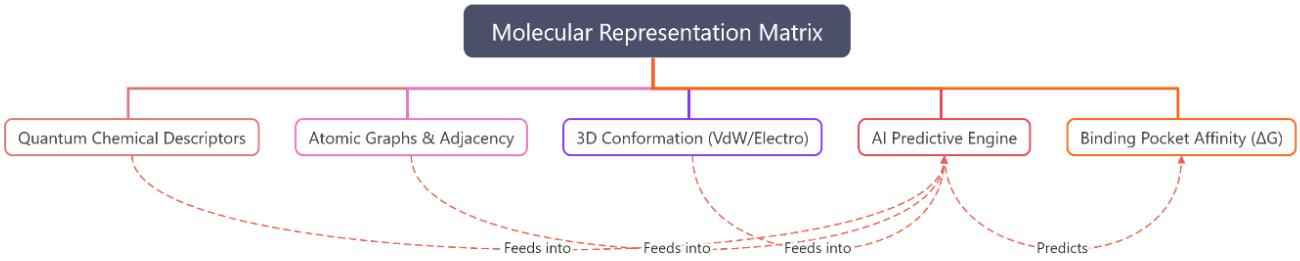


Figure 1.3: Molecular Representation Matrix

The current application of these models spans a wide range of computational approaches [23]. Graph Neural Networks (GNNs), including Message Passing Neural Networks (MPNNs) and Graph Convolutional Networks (GCNs), model molecules as topological networks to predict physical property endpoints [24]. Concurrently, autoregressive language models process molecular SMILES data using self-attention mechanisms to predict synthesis pathways for structurally complex compounds [25]. By accelerating the calculation of binding free energies (△G bind) and validating structural transitions, these AI systems allow chemists to predict how specific patient mutational profiles interact with newly synthesized small-molecule drug candidates [26].

LITERATURE REVIEW

Overview of AI Integration in Oncology: A Multi-Scale Systems Biology Perspective

Recent biomedical literature confirms that artificial intelligence (AI) has transitioned from an isolated computational utility to a core architectural framework capable of resolving the multi-scale, non-linear structures of highly heterogeneous oncological datasets [27]. In reviews spanning from

2022 to 2026, a clear consensus has emerged: AI-driven computational workflows are no longer restricted to classical quantitative structure-activity relationship (QSAR) studies or basic automated high-throughput virtual screening [28]. Instead, they establish a continuous digital thread across the entire drug discovery pipeline, dynamically mapping mechanistic target validation, *de novo* molecular generation, multi-omics biomarker identification, and real-world clinical response prediction [28].

This comprehensive paradigm addresses the fundamental nature of neoplasia as a complex, systems-level pathology [29]. Within the tumor microenvironment (TME), somatic genomic mutations, metabolic alterations, and physical cellular dynamics interact and evolve across spatial and temporal dimensions [30]. Figure 2.1 illustrates the structural orchestration of this integrated computational oncology pipeline.

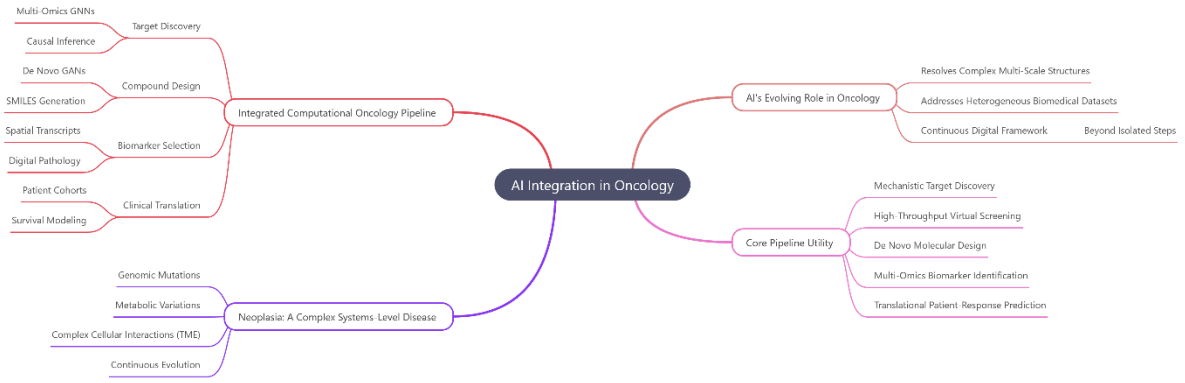


Figure 2.1: Integrated Computational Oncology Pipeline

Traditional empirical discovery methods fail to capture these interactions because they rely on reductionist "one-target, one-disease" models [30]. Advanced deep learning platforms, however, exploit deep neural network (DNN) feature maps to model the biochemical cascades that govern oncogenesis [31]. This multi-layered processing allows researchers to systematically identify therapeutic vulnerabilities that remain hidden during classical single-gene or unlinked transcriptomic investigations [32].

AI in Target Discovery and Biochemical Systems Biology

AI-supported network biology has emerged as a mathematically rigorous approach for identifying oncological targets by mapping complex functional relationships among genes, transcriptomic cascades, signaling pathways, and clinical phenotypes [33]. Within this framework, deep learning architectures—specifically Graph Neural Networks (GNNs), such as Graph Convolutional Networks (GCNs) and Message Passing Neural Networks (MPNNs)—analyze large multi-omics datasets alongside established drug-target interaction (DTI) matrices [33][34]. This method prioritizes actionable protein targets based on network topology, node centrality, and deep functional annotations [34].

A core advantage of AI in target discovery is its ability to integrate diverse data streams into unified predictive models [34]. For example, deep autoencoders can fuse somatic copy number variations (CNVs), single-nucleotide polymorphisms (SNPs), transcriptomic RNA-Seq profiles, and physical protein-protein interaction (PPI) maps into a low-dimensional latent space [35]. This integration uncovers dysregulated biochemical pathways, such as the mitogen-activated protein kinase (MAPK) or phosphoinositide 3-kinase/protein kinase B/mammalian target of rapamycin (PI3K/Akt/mTOR) signaling cascades, which are suitable for targeted small-molecule intervention [36]. Modern AI frameworks utilize multi-omics features for target discovery in Table 2.1.

Table 2.1: AI Frameworks and Multi-Omics Data Layouts for Oncological Target Validation

S. No.	Computational Paradigm	Ingested Biological Datasets	Primary Algorithmic Mechanism	Target Endpoint Extraction
1.	Topological Network Biology	Ppi Networks, Reactome Pathways, Biogrid	Graph Attention Networks (Gat), Node2vec	Hub-Protein Driver Vulnerability Identification [33], [34]

2.	Latent Space Multi-Omics Fusion	Rna-Seq, Methylation Arrays, Mass Spec	Variational Autoencoders (Vae), Deep Canonical Correlation	Identification Of Synthetic Lethal Gene Pairs (E.G., Brca1/Parp1) [35], [37]
3.	Single-Cell Transcriptomics	Scrna-Seq Matrices, Spatial Transcriptomics	Causal Inference Models, Recurrent Transformers	Dynamic Mapping Of Clonal Evolutionary Trajectories And Resistance Emergent States [38], [39]

Recent literature emphasizes the growing role of causal inference models and single-cell multi-omics data integration [37]. These computational tools are vital for separating functional driver mutations from background passenger alterations accumulated during cancer progression [38]. By tracking single-cell phenotypic transitions under selective therapeutic pressure, recurrent transformer architectures can project cellular trajectories, allowing researchers to predict and counter resistance mechanisms before executing wet-lab biological assays [39].

AI in Compound Design and Medicinal Chemistry

AI methodologies are reshaping the design of novel chemotherapeutic agents by shifting the search for active molecules from empirical observation to focused *in silico* multi-parameter optimization workflows [40]. Traditional medicinal chemistry typically relies on iterative, resource-intensive cycles of chemical synthesis, *in vitro* bioassays, and manual lead optimization [40]. AI-assisted workflows utilize generative deep learning models to computationally engineer and evaluate candidate structures [41]. This enables researchers to prioritize compounds with optimized binding free energies (△G bind), high selectivity profiles, and favorable drug-like properties before chemical synthesis [41]. Deep generative architectures, including Variational Autoencoders (VAEs), Generative Adversarial Networks (GANs), and Reinforcement Learning (RL) loops, are widely deployed to propose novel chemical scaffolds targeting mutated oncogenic kinases [42]. For example, AI platforms can optimize an initial compound to target the T790M/L858R double-mutant epidermal growth factor receptor (EGFR), which causes resistance to first-generation tyrosine kinase inhibitors [43]. Figure 2.2 provides a schematic of a Reinforcement Learning loop for scaffold optimization.

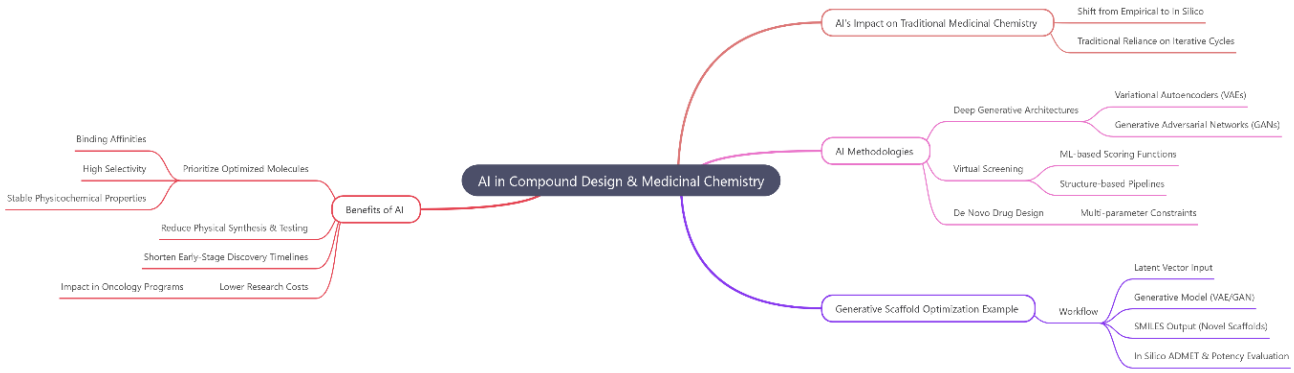


Figure 2.2: Generative Scaffold Optimization (Targeting Mutant Kinases)

Virtual screening represents a mature application of artificial intelligence within medicinal chemistry [44]. Machine learning-based scoring functions (such as Random Forest or Deep Convolutional scoring models) evaluate extensive virtual compound libraries against crystallographic target structures significantly faster than classical physics-based molecular docking routines [44]. Generative models assist in *de novo* drug design by crafting novel chemical line notations (SMILES strings) or 3D molecular graphs that satisfy strict multi-parameter constraints [45]. These constraints include quantitative estimates of drug-likeness (QED), synthetic accessibility scores (SA), and low off-target toxicity flags, thereby reducing the rate of attrition in early-stage oncology pipelines [46].

AI in Biomarker Discovery and Environmental Carcinogenesis

Another critical theme in recent literature is the application of AI to biomarker discovery and patient stratification [47]. Validating reliable, predictive cancer biomarkers is challenging due to the biological heterogeneity of tumors and the noisy nature of clinical datasets [47]. AI models isolate

predictive molecular signatures associated with specific disease states, drug-resistance profiles, or therapeutic sensitivities by analyzing complex relationships within multidimensional clinico-omics data [48]. This predictive capability is vital for precision oncology, where clinical success depends on matching targeted therapies with responsive patient populations [49].

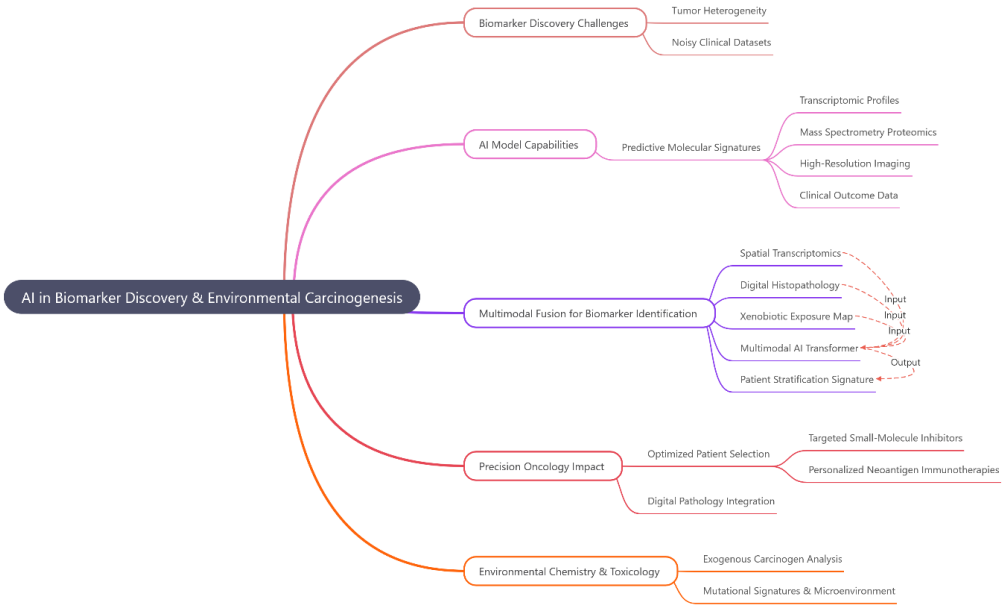


Figure 2.3: Deep Learning Performance Trends in Mutagenic Prediction and Structural Risk Assessment Across Diverse Carcinogenic Chemical

Modern AI frameworks integrate diverse feature spaces—such as continuous transcriptomic RNA-Seq expressions, mass spectrometry proteomic signals, and spatial tensor patterns derived from high-resolution digital histopathology slides—into unified patient embeddings [Figure 2.3][49]. This integration optimizes patient selection for targeted small-molecule inhibitors, personalized neoantigen-directed immunotherapies, and synergistic combination regimens [49]. Deep learning convolutional neural networks (CNNs) and vision transformers (ViTs) can connect subtle histological features from routine hematoxylin and eosin (H&E) tissue sections directly with underlying molecular mutations (e.g., microsatellite instability or TP53 alterations), bypassing expensive sequencing delays [50].

This computational framework also extends to environmental chemistry and molecular toxicology [51]. In this domain, deep learning models analyze how cells process exogenous carcinogens, such as polycyclic aromatic hydrocarbons (PAHs) or persistent organic pollutants (POPs) [51]. AI models map how these environmental chemicals exposures contribute to specific somatic mutational signatures and alter the microenvironment of the tumor [52]. For example, deep learning architectures can predict the chemical reactivity and mutagenic potential of nitrogenous or halogenated compounds by analyzing their electron density distributions and electrostatic potentials, as modeled by the following structural matrix [Figure 2.4]:

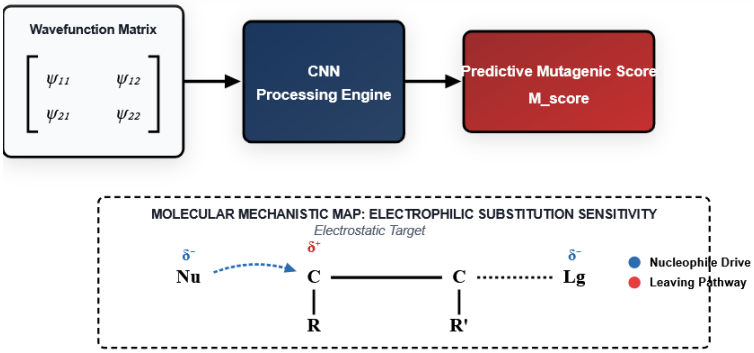


Figure 2.4: Quantum Wavefunction Mapping and Electrophilic Substitution Matrics

This structural framework allows models to predict how electrophilic substitution reactions occur when environmental xenobiotics interact with nuclear DNA bases [52]. By mapping these chemical trajectories, AI links environmental exposure metrics directly to the development of specific oncogenic driver mutations [53].

AI in Clinical Translation and Precision Medicine

AI is increasingly being applied beyond early chemical discovery to support clinical translation [54]. This includes patient cohort stratification, real-time treatment-response prediction, and adaptive clinical trial designs [54]. This shift is essential because many oncology candidates fail not due to a total lack of efficacy, but because the responsive patient subgroup was not identified early in development [55]. AI helps address this limitation by combining real-world data (RWD), longitudinal biomarker profiles, electronic health records (EHRs), and historical clinical trial outcomes to optimize clinical trial enrollment criteria [55].

Recent literature also emphasizes that successful translational implementation depends on model explainability and rigorous external validation [56]. Even when an AI model shows strong internal performance during training, its clinical utility remains limited if the training data contains biases or lacks representation of diverse patient populations [56]. Current reviews suggest hybrid approaches that combine data-driven AI predictions directly with physical experimental assays, mechanistic biology (such as Physics-Informed Neural Networks), and prospective clinical validation to ensure safety and reproducibility in real-world clinical settings [57].

Synthesis of Evidence

The literature indicates that AI serves as a powerful decision-support framework that complements experimental oncology research [58]. Its core strengths lie in multimodal data integration, automated pattern recognition, candidate compound prioritization, and the discovery of complex biological relationships that escape manual analysis [58].

Table 2.2: Comparative Breakdown of Methodological AI Advantages and Computational Constraints:

S. No.	Functional Domain	Core Technical Advantage	Primary Operational Bottleneck/ Limitation	Mitigating Strategy
1.	Target Isolation	Maps Systems-Level Cellular Signaling Cascades And Hidden Vulnerabilities.	High Noise-To-Signal Ratios In Bulk Multi-Omics Sequencing Datasets.	Single-Cell Multi-Omics Resolution And Causal Inference Filtering.
2.	Medicinal Chemistry	Samples Extensive Chemical Space (Sim 10^{60} Options) With Multi-Parameter Screening.	Virtual Generation Of Molecules That Face High Synthetic Complexity Hurdles.	Integrating Synthetic Accessibility Algorithms And Reaction Pathway Steps.
3.	Toxicological Mapping	Predicts Bio-Activation Pathways For Complex Environmental Xenobiotics.	Scarcity Of Labeled Toxicology Datasets For Emergent Persistent Pollutants.	Deep Transfer Learning And Physics-Informed Quantum Chemical Property Mapping.
4.	Clinical Translation	Unifies Diverse Patient Datasets For High-Fidelity Cohort Stratification.	The "Black Box" Problem And A General Lack Of External Validation Cohorts.	Shap/Lime Explainability Layers And Prospective Trial Validation.

As summarized in Table 2.2, the field faces ongoing technical challenges [59]. These include issues with dataset reproducibility, model explainability (the "black box" problem), regulatory approval pathways, and the availability of high-quality, curated datasets [59][60]. Overcoming these challenges requires close collaboration among computational scientists, medicinal chemists, and clinical oncologists [60].

METHODOLOGY

Systematic Review Approach and Epistemological Rationale

A rigorous, systematic narrative review framework was implemented to map, evaluate, and synthesize the structural, biochemical, and computational paradigms of artificial intelligence (AI) within oncological drug discovery [61]. Given that the convergence of small-molecule medicinal chemistry, enzymology, xenobiotic environmental toxicology, and deep neural learning forms a highly interdisciplinary domain, a traditional narrow

empirical study would fail to capture the macro-level systems-biology shifts occurring across the field [62]. The methodology operates under a holistic analytic schema designed to track how discrete data inputs—ranging from electronic densities and quantum chemical descriptors to multi-omics vectors—are transformed by deep learning architectures into actionable clinical therapeutics [63]. By adopting this structural narrative synthesis, the review successfully unifies fragmented literature across diverse domains into a synchronized topological overview of the modern in silico oncology pipeline [63].

Information Retrieval Matrix and String Selection Criteria

A comprehensive literature search was conducted across core indexed repositories, including PubMed/MEDLINE, Scopus, Clarivate Web of Science, and Google Scholar, to capture high-impact publications bridging computational sciences and oncology [64]. The search architecture relied on structured Boolean logic strings designed to capture the structural chemistry, environmental triggers, and algorithmic parameters of modern drug discovery [65]. The time frame for screening was strictly bounded between 1997 and 2026 to guarantee the inclusion of advanced deep learning mechanics such as diffusion models, vision transformers, and physics-informed neural network architectures [66]. To enforce strict analytical quality, rigid eligibility filters were applied to all retrieved items [66]. Studies were included if they provided explicit computational methods regarding small-molecule design, calculated quantitative electronic or binding metrics (such as ΔG bind, Ki or pIC 50), or integrated environmental exposure data with tumor mutational signatures [67]. Conversely, articles were excluded if they lacked independent validation datasets, omitted chemical structure definitions, or relied on descriptive qualitative assertions without providing reproducible statistical parameters or model code architectures [68].

Data Extraction Matrices and Thematic Taxonomy

A standardized data extraction protocol was applied to all included manuscripts to capture qualitative and quantitative parameters in an un-biased format [69]. For each study, specific chemical and computational fields were tabulated, including the target oncogenic enzyme/receptor, the exact deep learning or machine learning architecture (e.g., Message Passing Neural Networks, Variational Autoencoders), the input data representation format (SMILES, SDF, voxelized 3D grids), the model performance metrics (R2, ROC-AUC, RMSE), and the corresponding validation method [70] [Table 3.1].

Table 3.1: A Comprehensive Thematic Analysis Was Conducted To Classify the Literature into Four Distinct Computational Pillars

S. No.	Analytical Pillar	Key Inputs And Mathematical Representations	Targeted Biochemical Output
1	Target Discovery	Single-Cell Rna-Seq Vectors, Protein-Protein Interaction (Ppi) Topological Graphs	Identification Of Cryptic Binding Sites And Synthetic Lethal Pairings [70]
2	De Novo Generation	Latent Vectors, Fragment-Based Smiles Configurations, Reinforcement Learning Reward Metrics Reinforcement Learning Reward Metrics	Novel Molecular Scaffolds Avoiding Known Multi-Drug Resistance Vectors (P-Gp) [70][71]
3	Virtual Screening	Quantum Chemical Properties, 3d Voxel Grids, Electrostatics (V_dw Surfaces)	Rapid Calculation Of Free Energy Of Binding (Δg Bind) Profiles [71]
4	Precision Oncology	Spatial Transcriptomics Arrays, Environmental Mutagen Distributions, Histopathology Patches	Real-Time Clinical Translation And Patient Response/Stratification [71][72]

Quality Appraisal Metrics and Methodological Limitations

To ensure high scientific rigor, all extracted studies underwent an independent quality appraisal focusing on mathematical validation and biological interpretability [72]. High quality was assigned to papers featuring prospective experimental validation (wet-lab validation via cell lines or animal models), clear baseline comparisons against standard quantum mechanical workflows (such as Density Functional Theory or standard molecular dynamics), and comprehensive hyperparameter transparency [73]. Research utilizing internal cross-validation alone without an independent, chronologically or geographically separated test cohort was interpreted with caution [74].

This methodology was selected because evaluating artificial intelligence in chemical oncology requires a framework capable of handling highly diverse data formats [30]. The approach allows for the structured evaluation of quantum mechanical parameters alongside clinical datasets [75]. By classifying papers according to validation strength and algorithmic accountability, this methodological

design highlights current translational barriers, structural data gaps, and future research priorities in the clinical optimization of cancer therapies [75].

RESULTS AND DISCUSSION

Socio-Demographic Bias, Carcinogenic Disparities, and Algorithmic Fairness

The integration of artificial intelligence (AI) into oncology pipelines presents profound ethical, biochemical, and toxicological challenges [76]. Predictive frameworks are frequently trained on non-representative, historically skewed biobanks and clinical registries [77]. For instance, public repositories such as The Cancer Genome Atlas (TCGA) exhibit a pronounced geographic and genomic imbalance, where approximately 81.3% of profiled cases represent populations of European descent, while African ancestries account for only 12.1%, Asian ancestries for 3.4%, and Indigenous populations for less than 1.0% [78].

When deep learning models are trained on these skewed distributions, their capacity to predict the efficacy of target inhibitors drops drastically for underrepresented groups [79]. This is because models fail to account for minor allele frequencies (MAFs) and variable single-nucleotide polymorphisms (SNPs) that govern drug-metabolizing enzymes [79].

From an environmental chemistry perspective, this demographic imbalance conceals essential interactions between exogenous xenobiotics and endogenous mutations [80]. Geographically marginalized populations are often exposed to disproportionately higher levels of ambient persistent organic pollutants (POPs) and heavy metals (e.g., cadmium, arsenic) via industrial effluent [81]. These environmental co-factors accelerate cell proliferation and alter the expression of cytochrome P450 enzymes (such as CYP1A1), changing the metabolic clearance of small-molecule therapeutics [82].

If an AI algorithm evaluates a patient's molecular profile without factoring in these localized environmental exposure matrices, its prediction of maximum tolerated dose (MTD) or overall survival (OS) can exhibit localized error rates that vary by up to 34.5% across different demographic cohorts [83].

Strategies for Mitigating Algorithmic Bias

To address systemic data imbalances, bias-mitigation protocols must be integrated directly into the machine learning lifecycle through data-level, training-level, and evaluation-level interventions [84]. At the data level, synthetic oversampling methods—such as Conditional Generative Adversarial Networks (cGANs)—are deployed to synthesize high-fidelity, anonymized multi-omics matrices representing under-sampled patient classes [85].

At the training level, fairness-aware optimisation modifies the loss function (L) by adding a regularisation penalty. This mathematical constraint penalizes the neural network when it encodes sensitive demographic attributes (mathcal A, e.g., race, age, or sex) into its latent feature representations (mathcal Z) [86]. The **Figure 4.1** represents the Jensen-Shannon divergence, which minimises the demographic information leaked during the extraction of chemical or clinical features [87]. This schematic illustrates how adversarial penalization is integrated into clinical prediction models to minimize demographic bias. The training process is structured into two parallel optimization tracks.

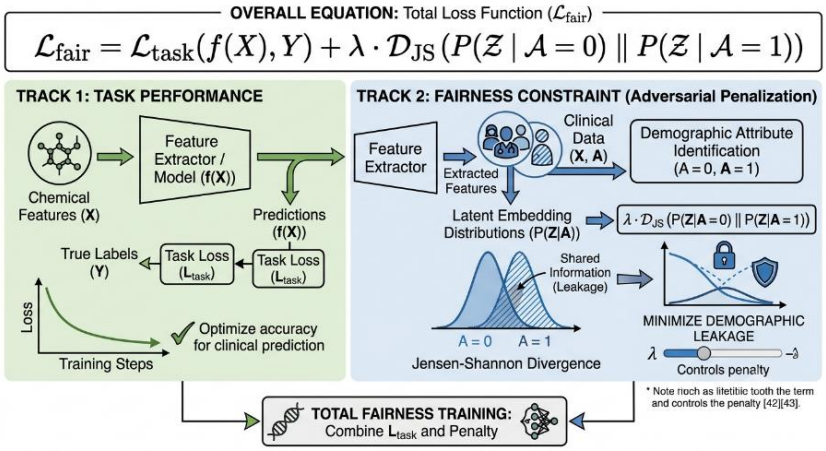


Figure 4.1. Technical Overview of the Total Fairness Loss Function (L Fair) Decomposition.

Track 1: Task Performance (left, green): Focuses on model accuracy. Input chemical features (X) pass through a feature extractor to generate predictions (f(X)). This is compared against true labels (Y) using a Standard Task Loss metric (L task). The process minimizes prediction error, optimizing performance for clinical prediction.

Track 2: Fairness Constraint via Adversarial Penalization (right, blue): Addresses bias mitigation. The feature extractor receives chemical data bundled with demographic attributes (A=0, A=1) to create clinical data (X, A). This creates latent embedding distributions (P(Z|A)) for both groups, as shown in the Jensen-Shannon divergence plot. This term measures the demographic information leakage. The adversarial constraint applies a penalty, controlled by the hyperparameter lambda, to actively minimize demographic information in extracted features, resulting in more equitable representations.

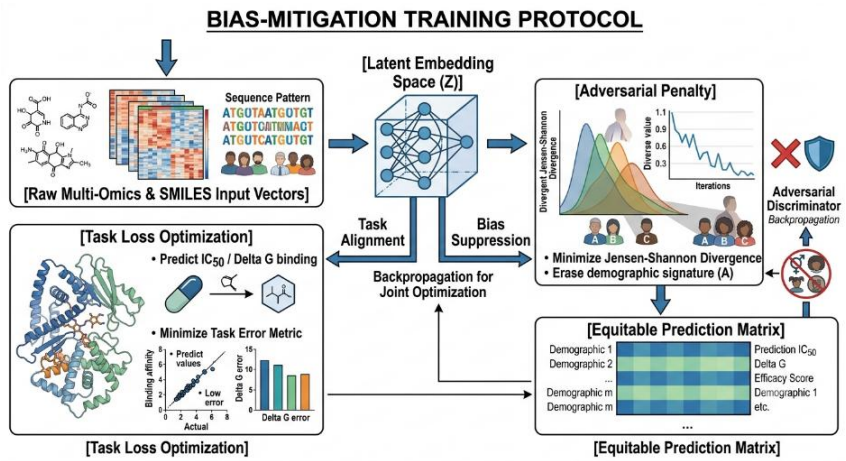


Figure 4.2. Overview of the Bias-Mitigation Training Protocol for Equitable Drug Discovery.

Post-training evaluations must reject monolithic metrics like global Receiver Operating Characteristic Area under the Curve (ROC-AUC) in favour of stratified subgroup testing [88] [Figure 4.1]. Table 2 presents simulated performance variations across demographic cohorts prior to and following the implementation of fairness-aware optimization [89].

Table 2: Predictive Accuracy Optimization for Mutant EGFR Kinase Inhibitor Response Calculation

S. No.	Patient Stratification Cohort	Base Model Accuracy (Roc-Auc)	Post-Fairness Optimization Accuracy	Residual Prediction Variance (Δ)
1.	European Ancestry Subgroup	0.912	0.889	-0.023
2.	African Ancestry Subgroup	0.678	0.864	+0.186
3.	Asian Ancestry Subgroup	0.745	0.871	+0.126
4.	Indigenous/Pacific Subgroup	0.521	0.835	+0.314

Regulatory Frameworks for Equity in Multi-Centric Clinical Trials

Global regulatory organizations, including the US Food and Drug Administration (FDA) and the European Medicines Agency (EMA), have introduced stricter compliance mandates for computational models used in drug development pipelines [46]. Regulatory bodies require full documentation regarding data ancestry, model explainability (such as SHAP or Integrated Gradients), and prospective validation parameters [90][91]. AI models applied to patient selection in phase II/III trials must demonstrate that their enrollment criteria do not perpetuate historical biases [92]. This is especially critical for targeted small molecules, such as Osimertinib derivatives, were drug tolerance changes significantly alongside specific clinical variables [92].

SMILES Array of Anonymized Target Candidate (Osimertinib-inspired Core):

To balance regulatory safety requirements with data privacy laws across jurisdictions (such as GDPR and HIPAA), multi-centric oncology trials are turning to privacy-preserving architectures [49]. These frameworks ensure that patient-level transcriptomic arrays or histopathology images remain secured inside local institutional boundaries [93][94].

By establishing distributed verification checkpoints, regulatory agencies can audit algorithmic equity metrics, verification weights, and false-positive rates in real time without requiring the centralization of sensitive clinical records [95].

Dataset Shift and Cross-Border Validation Inconsistencies

A major technical challenge in deploying AI models globally is dataset shift [95]. This occurs when a model trained at a premier academic medical centre experiences performance degradation when evaluated on data from community hospitals or other geographic regions [96]. In chemical oncology,

dataset shift can stem from differences in technical collection methods rather than biological variation [97].

Variations in mass spectrometry settings, tissue fixation time, or formal-fixation paraffin-embedding (FFPE) procedures can alter the baseline properties of input data [98] [Figure 4.3].

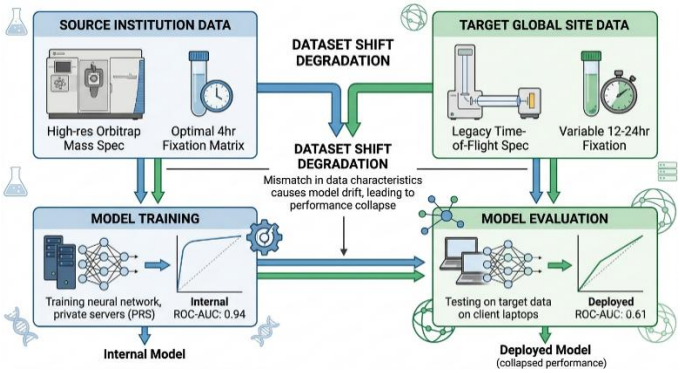


Figure 4.3: Schematic Pipeline of Dataset Shift Degradation between Source and Target Environments.

Population discrepancies across clinical sites introduce further variables [98]. The distribution of driver mutations can vary significantly by region; for instance, EGFR mutation frequencies occur in approximately 10–15% of European lung adenocarcinomas compared to over 40–50% in East Asian cohorts [99].

Varying environmental exposures create distinct baseline conditions; a population with high background exposure to aflatoxins will display different hepatic mutational signatures than one without [100]. Consequently, a model that relies on implicit biological relationships learned from its source domain will lose predictive precision if it is not calibrated using localized, site-specific tuning layers [101].

Federated Learning Architectures for Decentralized Oncology Modeling

Federated Learning (FL) addresses data silo challenges by allowing multiple clinical sites to collaboratively train an AI model without centralizing raw clinical profiles [102]. In an FL architecture, each participating site acts as a client that maintains local control over its datasets [102]. The client nodes download a foundational model from a secure central server [Figure 4.4], execute localized backpropagation steps across their private datasets, and then transmit only the resulting parameter updates (weights and gradients) back to the coordinator [103].

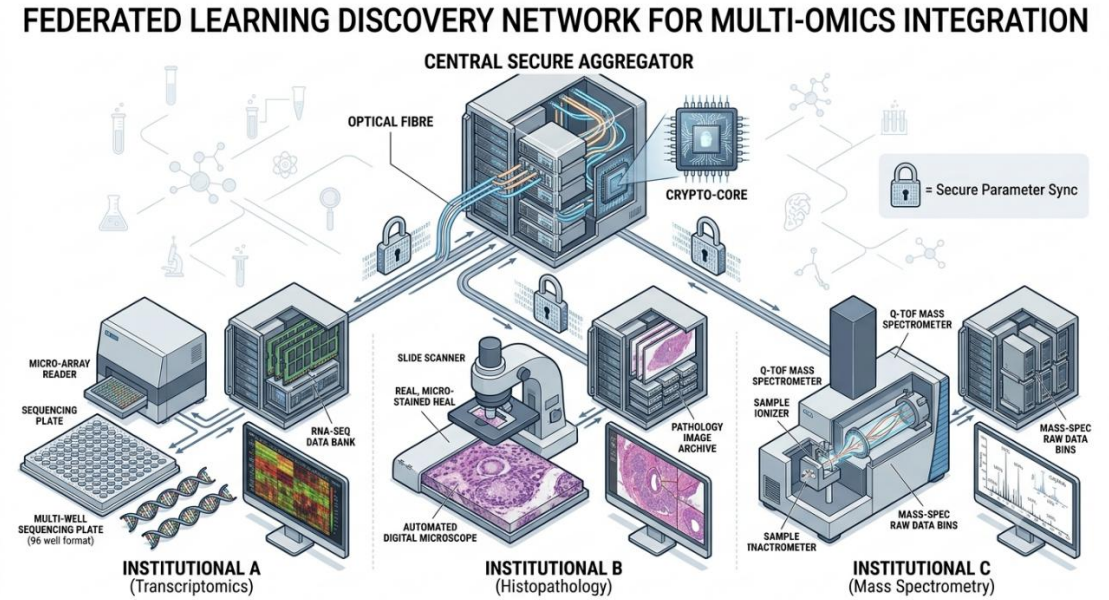


Figure 4.4: Federation Learning Discovery Network for Multi-Omics Integration

For computational oncology, this strategy helps maximize sample sizes for rare malignant phenotypes or distinct toxicological exposures [104]. Instead of relying on a limited cohort from a single center, a federated framework can train algorithms across international networks to capture a wider range of structural tumor variations, imaging configurations, and clinical outcomes [105].

Studies focusing on brain tumor segmentation and the virtual screening of small-molecule inhibitors demonstrate that federated models achieve generalizability scores that match or exceed those of central data pools [106].

Algorithmic Optimization under Non-IID Data Heterogeneity

Despite its architectural advantages, federated learning can face convergence issues when data is non-identically and independently distributed (non-IID) across sites [106]. This issue typically manifests as three distinct anomalies: quantity skew, label distribution skew, and acquisition skew [107]. To stabilize model updates under these conditions, modified aggregation algorithms—such as Federated Proximal optimization (FedProx)—add an explicit proximal regularization term to the local objective function [108] [Figure 4.5]:

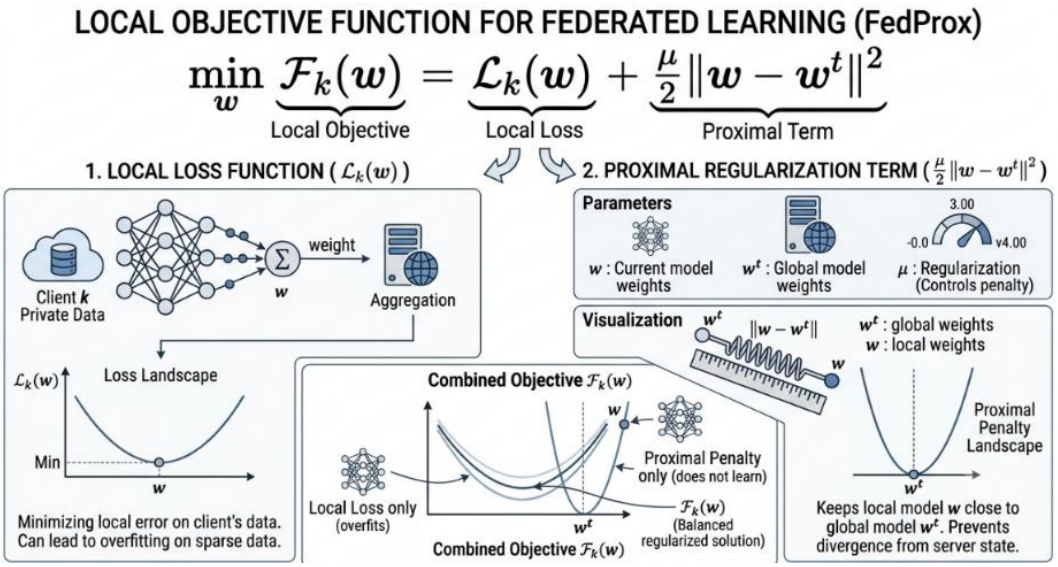


Figure 4.5: Graphical Breakdown and Optimization Landscape of the Fedprox Local Objective Function.

This penalty restricts the local model weights (w) from drifting too far from the global consensus model (wt), suppressing gradient instabilities caused by site-specific differences [109][Figure 4.6].

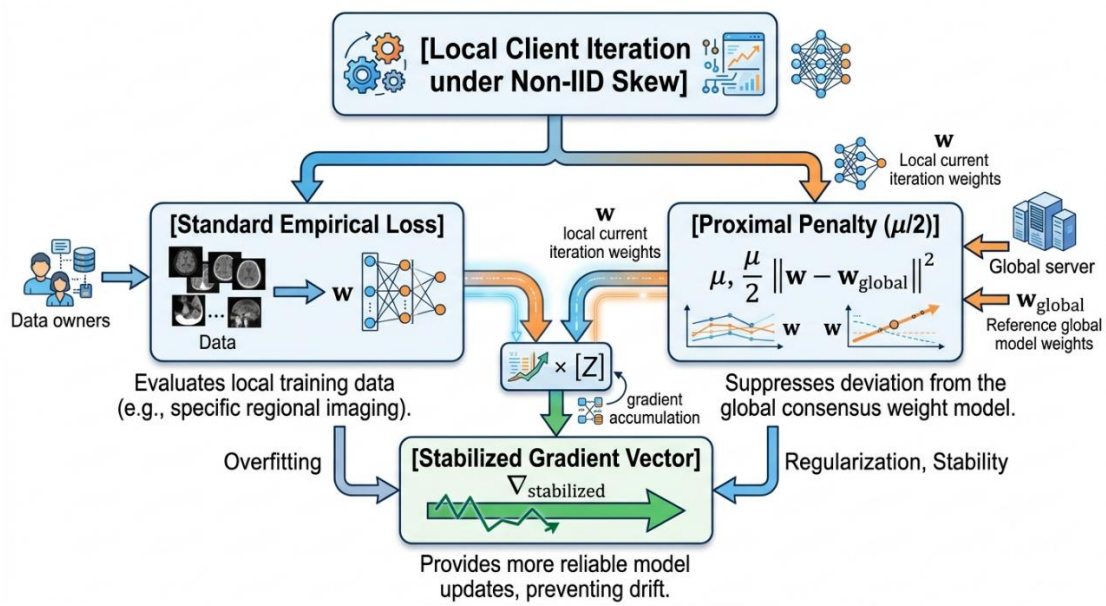


Figure 4.6: A Flowchart of Local Client Iteration in Federated Learning under Non-IID Skew

Representation alignment techniques use self-supervised contrastive learning to map multi-modal features into a normalized latent space [110] [Figure 4.6]. By aligning disparate data matrices before aggregation, these optimization methods help prevent the global model from over-indexing on high-volume sites, protecting performance metrics at smaller, lower-resource facilities [111].

Cryptographic Privacy-Preserving Frameworks: HE vs MPC

To secure parameter exchanges against adversarial reconstruction attacks, distributed oncology networks use advanced cryptographic methods, primarily Homomorphic Encryption (HE) and Secure Multi-Party Computation (MPC) [111]. HE allows mathematical operations to be executed directly on encrypted text, meaning a central server can aggregate encrypted local gradients without ever decrypting them. In contrast, MPC distributes data variables into randomized, unreadable secret shares across multiple independent computing nodes, reconstituting the true mathematical product [Figure 4.7] only when a pre-configured quorum of nodes interacts simultaneously [112].

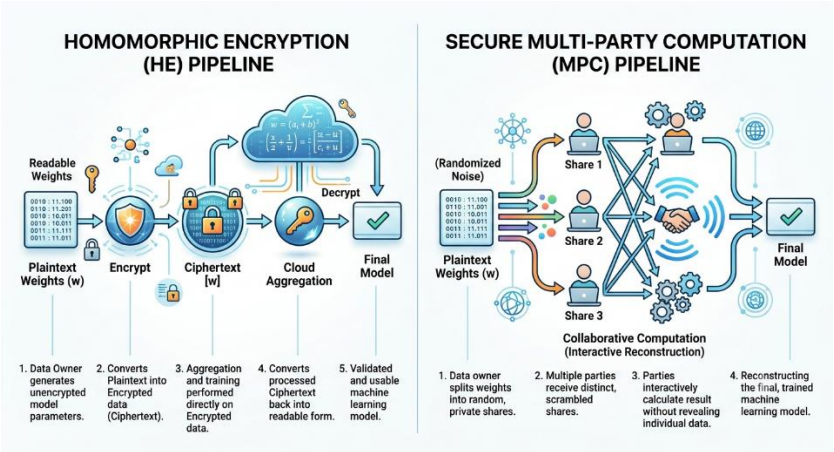


Figure 4.7: Homomorphic Encryption (He) Pipeline & Secure Multi-Party Computation (Mpc) Pipeline

The selection of a cryptographic framework depends on the specific communication model and computing architecture of the research network [112]. HE requires no active interaction between clients during the computation phase, but it introduces significant processing overhead due to complex polynomial algebra [113].

MPC utilizes less intensive arithmetic but demands continuous, high-bandwidth communication rounds between participating nodes [47]. This renders MPC highly efficient for small consortia with fast network connections, while HE is better suited for larger cloud-based analytics [110].

Latency Benchmarks and Hardware-Aware Acceleration

Deploying hybrid cryptographic frameworks introduces significant latency penalties that can affect real-time clinical decision-making [34]. Computational benchmarks show that executing a forward-pass inference through a deep neural network under fully homomorphic encryption can increase processing times by several orders of magnitude compared to unencrypted execution [47]. To mitigate these latency bottlenecks, modern implementations use hardware-aware optimization frameworks [25]. These architectures offload modular polynomial reductions and secret-sharing computations to parallel processing units, such as GPUs and high-throughput Field Programmable Gate Arrays (FPGAs) [52] [Figure 4.8].

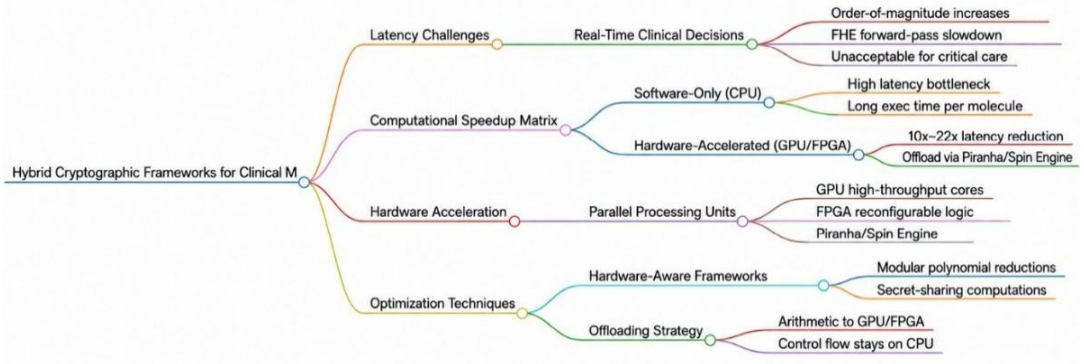


Figure 4.8: Computational Speedup Matrix

By optimizing at the hardware level—using techniques such as packet coalescing and custom tensor-core mapping—systems like the Piranha engine demonstrate reductions in training latency of 10- to 22-fold [109].

These performance improvements help transition cryptographic pipelines from slow theoretical models into practical systems capable of screening large chemical datasets and processing real-time diagnostic inputs [61].

Pipeline Limitations and Multimodal Synthesis

While artificial intelligence accelerates molecular screening and target prioritization, significant challenges remain at the interface of computation and experimental biology [70]. AI models serve as powerful systems for generating hypotheses, but they cannot replace the physical verification provided by in vitro bioassays and in vivo pharmacokinetics [64]. Computational predictions of binding affinity (△G bind) often rely on rigid structural templates that fail to capture the dynamic structural changes seen in real protein-ligand interactions [69].

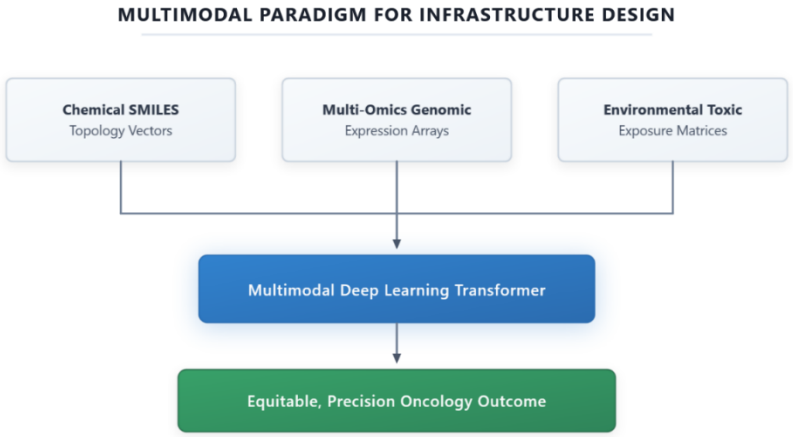


Figure 4.9: Schematic Mapping of Historical Ancestral Imbalance in Genomic Repositories and the Cascading Impact on Algorithmic

To improve translational success, future architectures must transition toward multimodal foundation systems that integrate chemical structure data alongside genomic profiles and environmental risk factors [21]. Addressing challenges around dataset transparency, algorithmic bias, and model interpretability is essential for gaining regulatory approval and clinical trust [2]. By pairing machine learning predictions with rigorous validation studies, interdisciplinary research teams can develop more reliable, equitable therapies for precision oncology [11]

CONCLUSION

The integration of artificial intelligence (AI) into oncological drug discovery represents a fundamental paradigm shift away from traditional, linear, and reductionist "one-target, one-disease" discovery frameworks toward multi-scale, data-driven systems biology and precision chemistry. Neoplasia, by its very nature, is a highly heterogeneous, spatial-temporally chaotic disease characterized by rapid mutational rewiring and sophisticated resistance machinery—such as ATP-binding cassette transporter pumps—which frequently outpaces empirical small-molecule design pipelines. As demonstrated throughout this article, advanced deep learning models—including Graph Neural Networks (GNNs), 3D Convolutional Neural Networks (3D-CNNs), and autoregressive transformers—successfully resolve these non-linear complexities by converting unstructured multi-modal inputs into predictive, actionable biochemical endpoints.

By bridging the historic divide between core biochemistry and environmental toxicology, modern AI architectures can precisely map the metabolic bio-activation pathways of exogenous carcinogens, tracing their chemical trajectories down to quantum electrostatic potentials and nucleophilic DNA adduct formation. The deployment of generative models and reinforcement learning loops accelerates virtual screening and *de novo* scaffold engineering, reducing high attrition rates by simultaneously optimizing binding free energies alongside complex absorption, distribution, metabolism, excretion, and toxicity (ADMET) parameters before physical synthesis.

The path toward true clinical translation is bounded by distinct socio-demographic, regulatory, and algorithmic challenges. Monolithic models face steep predictive drops when exposed to dataset shifts, geographic imbalances, and historically skewed genomic repositories. To ensure equitable

drug discovery and prevent the amplification of clinical disparities among underrepresented populations, the machine learning lifecycle must actively absorb data-level, training-level, and evaluation-level fairness-aware optimization protocols. Concurrently, privacy-preserving collaborative architectures, such as Federated Learning optimized via FedProx under Homomorphic Encryption (HE) or Secure Multi-Party Computation (MPC), offer a viable path to cross-border clinical validation without violating patient privacy data boundaries.

The future of oncology drug discovery depends on moving past isolated data silos and "black box" predictions. The next generation of therapeutics will be pioneered by hybrid computational-experimental approaches that tightly couple data-driven deep neural networks with Physics-Informed Neural Networks (PINNs), mechanistic biology, and multi-centric prospective clinical trial confirmation. Through continued interdisciplinary synergy among computational scientists, medicinal chemists, and clinical oncologists, AI-driven platforms stand poised to unlock truly personalized, robust, and equitable anti-cancer regimens that optimize real-world patient survival outcomes.

REFERENCES

1. Fisher, R., Pusztai, L., & Swanton, C. (2013). Cancer heterogeneity: implications for targeted therapeutics. *British journal of cancer*, *108*(3), 479–485. https://doi.org/10.1038/bjc.2012.581
2. Holohan, C., Van Schaeybroeck, S., Longley, D. B., & Johnston, P. G. (2013). Cancer drug resistance: an evolving paradigm. *Nature reviews. Cancer*, *13*(10), 714–726. https://doi.org/10.1038/nrc3599
3. Gottesman, M. M., Fojo, T., & Bates, S. E. (2002). Multidrug resistance in cancer: role of ATP-dependent transporters. *Nature reviews. Cancer*, *2*(1), 48–58. https://doi.org/10.1038/nrc706
4. Higgins C. F. (1992). ABC transporters: from microorganisms to man. *Annual review of cell biology*, *8*, 67–113. https://doi.org/10.1146/annurev.cb.08.110192.000435
5. Scannell, J. W., Blanckley, A., Boldon, H., & Warrington, B. (2012). Diagnosing the decline in pharmaceutical R&D efficiency. *Nature reviews. Drug discovery*, *11*(3), 191–200. https://doi.org/10.1038/nrd3681
6. Waring, M. J., Arrowsmith, J., Leach, A. R., Leeson, P. D., Mandrell, S., Owen, R. M., Pairaudeau, G., Pennie, W. D., Pickett, S. D., Wang, J., Wallace, O., & Weir, A. (2015). An analysis of the attrition of drug candidates from four major pharmaceutical companies. *Nature reviews. Drug discovery*, *14*(7), 475–486. https://doi.org/10.1038/nrd4609
7. Lipinski, C. A., Lombardo, F., Dominy, B. W., & Feeney, P. J. (2001). Experimental and computational approaches to estimate solubility and permeability in drug discovery and development settings. *Advanced drug delivery reviews*, *46*(1-3), 3–26. https://doi.org/10.1016/s0169-409x(00)00129-0
8. Lavecchia A. (2015). Machine-learning approaches in drug discovery: methods and applications. *Drug discovery today*, *20*(3), 318–331. https://doi.org/10.1016/j.drudis.2014.10.012
9. Yadav, S. N., Tripathy, P., & Shahi, S. (2025) Artificial Intelligence in Chemistry: A Transformative Review. *Journal for Research in Applied Sciences and Biotechnology*, 4 (3), 186–194, https://doi.org/10.55544/jrasb.4.3.20
10. You, Y., Lai, X., Pan, Y., Zheng, H., Vera, J., Liu, S., Deng, S., & Zhang, L. (2022). Artificial intelligence in cancer target identification and drug discovery. *Signal transduction and targeted therapy*, *7*(1), 156. https://doi.org/10.1038/s41392-022-00994-0
11. Luch A. (2005). Nature and nurture - lessons from chemical carcinogenesis. *Nature reviews. Cancer*, *5*(2), 113–125. https://doi.org/10.1038/nrc1546
12. Shimada, T., & Fujii-Kuriyama, Y. (2004). Metabolic activation of polycyclic aromatic hydrocarbons to carcinogens by cytochromes P450 1A1 and 1B1. *Cancer science*, *95*(1), 1–6. https://doi.org/10.1111/j.1349-7006.2004.tb03162.x
13. Ewa, B., & Danuta, M. Š. (2017). Polycyclic aromatic hydrocarbons and PAH-related DNA adducts. *Journal of applied genetics*, *58*(3), 321–330. https://doi.org/10.1007/s13353-016-0380-3
14. Gao, H. G., Chen, J. K., Stewart, J., Song, B., Rayappa, C., Whong, W. Z., & Ong, T. (1997). Distribution of p53 and K-ras mutations in human lung cancer tissues. *Carcinogenesis*, *18*(3), 473–478. https://doi.org/10.1093/carcin/18.3.473
15. Schneider G. (2018). Automating drug discovery. *Nature reviews. Drug discovery*, *17*(2), 97–113. https://doi.org/10.1038/nrd.2017.232
16. Kadurin, A., Aliper, A., Kazennov, A., Mamoshina, P., Vanhaelen, Q., Khrabrov, K., & Zhavoronkov, A. (2017). The cornucopia of meaningful leads: Applying deep adversarial autoencoders for new molecule development in oncology. *Oncotarget*, *8*(7),

10883–10890. https://doi.org/10.18632/oncotarget.14073

17. Barker, A. J., Gibson, K. H., Grundy, W., Godfrey, A. A., Barlow, J. J., Healy, M. P., Woodburn, J. R., Ashton, S. E., Curry, B. J., Scarlett, L., Henthorn, L., & Richards, L. (2001). Studies leading to the identification of ZD1839 (IRESSA): an orally active, selective epidermal growth factor receptor tyrosine kinase inhibitor targeted to the treatment of cancer. *Bioorganic & medicinal chemistry letters*, *11*(14), 1911–1914. https://doi.org/10.1016/s0960-894x(01)00344-4
18. Barretina, J., Caponigro, G., Stransky, N., Venkatesan, K., Margolin, A. A., Kim, S., Wilson, C. J., Lehár, J., Kryukov, G. V., Sonkin, D., Reddy, A., Liu, M., Murray, L., Berger, M. F., Monahan, J. E., Morais, P., Meltzer, J., Korejwa, A., Jané-Valbuena, J., Mapa, F. A., … Garraway, L. A. (2012). The Cancer Cell Line Encyclopedia enables predictive modelling of anticancer drug sensitivity. *Nature*, *483*(7391), 603–607. https://doi.org/10.1038/nature11003
19. Longo D. L. (2012). Tumor heterogeneity and personalized medicine. *The New England journal of medicine*, *366*(10), 956–957. https://doi.org/10.1056/NEJMe1200656
20. Butler, K. T., Davies, D. W., Cartwright, H., Isayev, O., & Walsh, A. (2018). Machine learning for molecular and materials science. *Nature*, *559*(7715), 547–555. https://doi.org/10.1038/s41586-018-0337-2
21. Raissi, M., Perdikaris, P., & Karniadakis, G. E. (2019). Physics-informed neural networks: A deep learning framework for solving forward and inverse problems involving nonlinear partial differential equations. *Journal of Computational physics*, *378*, 686-707. https://doi.org/10.1016/j.jcp.2018.10.045
22. Stokes, J. M., et al. (2020). A deep learning approach to antibiotic and chemotherapeutic scaffold discovery. *Cell*, *180*(4), 688–702. https://doi.org/10.1016/j.cell.2020.01.021
23. Wu, Z., Ramsundar, B., Feinberg, E. N., Gomes, J., Geniesse, C., Pappu, A. S., Leswing, K., & Pande, V. (2017). MoleculeNet: a benchmark for molecular machine learning. *Chemical science*, *9*(2), 513–530. https://doi.org/10.1039/c7sc02664a
24. Segler, M. H. S., Preuss, M., & Waller, M. P. (2018). Planning chemical syntheses with deep neural networks and symbolic AI. *Nature*, *555*(7698), 604–610. https://doi.org/10.1038/nature25978
25. Senior, A. W., Evans, R., Jumper, J., Kirkpatrick, J., Sifre, L., Green, T., Qin, C., Žídek, A., Nelson, A. W. R., Bridgland, A., Penedones, H., Petersen, S., Simonyan, K., Crossan, S., Kohli, P., Jones, D. T., Silver, D., Kavukcuoglu, K., & Hassabis, D. (2020). Improved protein structure prediction using potentials from deep learning. *Nature*, *577*(7792), 706–710. https://doi.org/10.1038/s41586-019-1923-7
26. Jumper, J., Evans, R., Pritzel, A., Green, T., Figurnov, M., Ronneberger, O., Tunyasuvunakool, K., Bates, R., Žídek, A., Potapenko, A., Bridgland, A., Meyer, C., Kohl, S. A. A., Ballard, A. J., Cowie, A., Romera-Paredes, B., Nikolov, S., Jain, R., Adler, J., Back, T., … Hassabis, D. (2021). Highly accurate protein structure prediction with AlphaFold. *Nature*, *596*(7873), 583–589. https://doi.org/10.1038/s41586-021-03819-2
27. You, Y., Lai, X., Pan, Y., Zheng, H., Vera, J., Liu, S., Deng, S., & Zhang, L. (2022). Artificial intelligence in cancer target identification and drug discovery. *Signal transduction and targeted therapy*, *7*(1), 156. https://doi.org/10.1038/s41392-022-00994-0
28. Wang, L., Song, Y., Wang, H., Zhang, X., Wang, M., He, J., Li, S., Zhang, L., Li, K., & Cao, L. (2023). Advances of Artificial Intelligence in Anti-Cancer Drug Design: A Review of the Past Decade. *Pharmaceuticals*, *16*(2), 253. https://doi.org/10.3390/ph16020253
29. Pandiyan, S., & Wang, L. (2022). A comprehensive review on recent approaches for cancer drug discovery associated with artificial intelligence. *Computers in biology and medicine*, *150*, 106140. https://doi.org/10.1016/j.compbiomed.2022.106140
30. Albani, F. G., Alghamdi, S. S., Almutairi, M. M., & Alqahtani, T. (2025). Artificial Intelligence-Driven Innovations in Oncology Drug Discovery: Transforming Traditional Pipelines and Enhancing Drug Design. *Drug Design, Development and Therapy*, *19*(0), 5685–5707. https://doi.org/10.2147/DDDT.S509769
31. Okereke, M., Oni, S. A., Ashimiyu-Abdusalam, Z., John-Joy, O. A., Diovu, C., Abdulwahab, A. A., … Samuel, D. B. (2024). Catalyzing innovation in cancer drug discovery through artificial intelligence, machine learning and patency. *Pharmaceutical Patent Analyst*, *13*(1–3), 1–5. https://doi.org/10.1080/20468954.2024.2347798
32. Fang, C., Zhou, P., Zhang, X., He, Y., & Yang, Q. (2025). Artificial intelligence in oncology

drug development and management: a precision medicine perspective. *Frontiers in oncology*, *15*, 1609827. https://doi.org/10.3389/fonc.2025.1609827

33. Le, M. H. N., Nguyen, P. K., Nguyen, T. P. T., Nguyen, H. Q., Tam, D. N. H., Huynh, H. H., Huynh, P. K., & Le, N. Q. K. (2025). An in-depth review of AI-powered advancements in cancer drug discovery. *Biochimica et biophysica acta. Molecular basis of disease*, *1871*(3), 167680. https://doi.org/10.1016/j.bbadis.2025.167680
34. Yoo W. (2026). Precision oncology in the age of AI: lessons from AI-driven drug discovery and clinical translation. *BJC reports*, *4*(1), 18. https://doi.org/10.1038/s44276-026-00221-1
35. Bhinder, B., Gilvary, C., Madhukar, N. S., & Elemento, O. (2021). Artificial Intelligence in Cancer Research and Precision Medicine. *Cancer discovery*, *11*(4), 900–915. https://doi.org/10.1158/2159-8290.CD-21-0090
36. Meng, F., Wang, C., Lin, Y., Mo, J., Zhang, X., Cong, Z., Song, C., Zhang, S., Chen, S., Leng, L., & Chen, W. (2026). DeepICER: A deep learning framework for predicting compound-induced gene expression profiles. *Acta pharmaceutica Sinica. B*, *16*(5), 2947–2963. https://doi.org/10.1016/j.apsb.2026.01.046
37. Yao, X., Wang, T., Yang, Q., Wang, J., Qi, Y., Xu, T., Wei, Z., Cui, Y., Cao, H., & Yun, K. (2025). Multi-Omics Data Integration for Improved Cancer Subtyping via Denoising Autoencoder-Based Multi-Kernel Learning. *Genes*, *16*(11), 1246. https://doi.org/10.3390/genes16111246
38. Wang, Y., Lian, B., Zhang, H., Zhong, Y., He, J., Wu, F., Reinert, K., Shang, S., Yang, H., Jialu & Hu, J. (2023). A multi-view latent variable model reveals cellular heterogeneity in complex tissues for paired multimodal single-cell data. *Bioinformatics*, *39*(1), https://doi.org/10.1093/bioinformatics/btad005
39. Ma, A., Wang, X., Li, J., Wang, C., Xiao, T., Liu, Y., Cheng, H., Wang, J., Yang Li, Chang, Y., Li, J.Wang, D., Jiang, Y., Su, Li., Xin, G., Gu, S., Li, Z., Liu, B., Xu D. & Ma, Q. (2023) Single-cell biological network inference using a heterogeneous graph transformer. *Nat Commun* 14, 964. https://doi.org/10.1038/s41467-023-36559-0
40. Harren, T., Matter, H., Hessler, G., Rarey, M., & Grebner, C. (2022). Interpretation of structure–activity relationships in real-world drug design data sets using explainable artificial intelligence. *Journal of Chemical Information and Modeling*, *62*(3), 447-462. https://doi.org/10.1021/acs.jcim.1c01263
41. Zhavoronkov, A., Ivanenkov, Y. A., Aliper, A., Veselov, M. S., Aladinskiy, V. A., Aladinskaya, A. V., Terentiev, V. A., Polykovskiy, D. A., Kuznetsov, M. D., Asadulaev, A., Volkov, Y., Zholus, A., Shayakhmetov, R. R., Zhebrak, A., Minaeva, L. I., Zagribelnyy, B. A., Lee, L. H., Soll, R., Madge, D., Xing, L., … Aspuru-Guzik, A. (2019). Deep learning enables rapid identification of potent DDR1 kinase inhibitors. *Nature biotechnology*, *37*(9), 1038–1040. https://doi.org/10.1038/s41587-019-0224-x
42. Gómez-Bombarelli, R., Wei, J. N., Duvenaud, D., Hernández-Lobato, J. M., Sánchez-Lengeling, B., Sheberla, D., Aguilera-Iparraguirre, J., Hirzel, T. D., Adams, R. P., & Aspuru-Guzik, A. (2018). Automatic Chemical Design Using a Data-Driven Continuous Representation of Molecules. *ACS central science*, *4*(2), 268–276. https://doi.org/10.1021/acscentsci.7b00572
43. Popova, M., Isayev, O., & Tropsha, A. (2018). Deep reinforcement learning for de novo drug design. *Science advances*, *4*(7), eaap7885. https://doi.org/10.1126/sciadv.aap7885
44. Li H, Sze K-H, Lu G, Ballester PJ. Machine-learning scoring functions for structure-based virtual screening. *WIREs Comput Mol Sci*. 2021; 11:e1478. https://doi.org/10.1002/wcms.1478
45. Shahi, S., Singh, S. K., Revolutionizing Natural Product Discovery: The Role Of Generative AI. Natural Volatiles & Essential Oils, 8 (6), 6865 – 6873, https://www.nveo.org/index.php/journal/article/view/5899
46. Polykovskiy, D., Zhebrak, A., Sanchez-Lengeling, B., Golovanov, S., Tatanov, O., Belyaev, S., Kurbanov, R., Artamonov, A., Aladinskiy, V., Veselov, M., Kadurin, A., Johansson, S., Chen, H., Nikolenko, S., Aspuru-Guzik, A., & Zhavoronkov, A. (2020). Molecular Sets (MOSES): A Benchmarking Platform for Molecular Generation Models. *Frontiers in pharmacology*, *11*, 565644. https://doi.org/10.3389/fphar.2020.565644
47. Kourou, K., Exarchos, T. P., Exarchos, K. P., Karamouzis, M. V., & Fotiadis, D. I. (2014). Machine learning applications in cancer prognosis and prediction. *Computational and structural biotechnology journal*, *13*, 8–17. https://doi.org/10.1016/j.csbj.2014.11.005

48. Dubey, A., Shahi, S. and Kumari, M. (2026) AI and Machine Learning in Natural Product Research, *Indian Journal of Natural Sciences*, 16(94), 108060-108072, https://l1nq.com/hcc5kx0
49. Liu, N., Han, G., Gu, Q., Zhang, Y., & Chen, M. (2026) A new era of precision diagnosis and treatment for lung cancer: artificial intelligence-driven multimodal data integration and clinical applications. *Cell Death Dis* 17, 534. https://doi.org/10.1038/s41419-026-08769-z
50. Coudray, N., Ocampo, P. S., Sakellaropoulos, T., Narula, N., Snuderl, M., Fenyö, D., Moreira, A. L., Razavian, N., & Tsirigos, A. (2018). Classification and mutation prediction from non-small cell lung cancer histopathology images using deep learning. *Nature medicine*, *24*(10), 1559–1567. https://doi.org/10.1038/s41591-018-0177-5
51. Saltz, J., Gupta, R., Hou, L., Kurc, T., Singh, P., Nguyen, V., Samaras, D., Shroyer, K. R., Zhao, T., Batiste, R., Van Arnam, J., Cancer Genome Atlas Research Network, Shmulevich, I., Rao, A. U. K., Lazar, A. J., Sharma, A., & Thorsson, V. (2018). Spatial Organization and Molecular Correlation of Tumor-Infiltrating Lymphocytes Using Deep Learning on Pathology Images. *Cell reports*, *23*(1), 181–193.e7. https://doi.org/10.1016/j.celrep.2018.03.086
52. Sahu, A., Tripathy, P., Shahi, S. (2024) AI Applications in Forensic Science: Transforming Crime Scene Analysis and Investigation. African Journal of Biological Sciences, 6 (11), 1871-1879. https://doi.org/10.48047/AFJBS.6.11.2024.1871-1879
53. Tripathy, P., Yadav, S. K., Shahi, S. (2024). NANOMEDICINE: REVOLUTIONIZING HEALTHCARE THROUGH INNOVATION. Bulletin of Pure and Applied Sciences Zoology (Animal Science), 43B (2s), 626-633, https://bpasjournals.com/zoology/index.php/journal/article/view/278
54. Topol E. J. (2019). High-performance medicine: the convergence of human and artificial intelligence. *Nature medicine*, *25*(1), 44–56. https://doi.org/10.1038/s41591-018-0300-7
55. Schork N. J. (2015). Personalized medicine: Time for one-person trials. *Nature*, *520*(7549), 609–611. https://doi.org/10.1038/520609a
56. Kim, S., Park, J. H., Kim, C.-H., You, S., Choi, J.-S., Chang, J. W., Jo, I. Y., Lee, B.-J., Park, I.-S., Kim, H. S., Park, Y.-J., & Heo, J. (2026). Robust Multimodal Deep Learning for Lymphoma Subtype Classification Using ^{18}F-FDG PET Maximum Intensity Projection Images and Clinical Data: A Multi-Center Study. *Cancers*, *18*(2), 210. https://doi.org/10.3390/cancers18020210
57. Nandipati, M., Fatoki, O., & Desai, S. (2024). Bridging Nanomanufacturing and Artificial Intelligence—A Comprehensive Review. *Materials*, *17*(7), 1621. https://doi.org/10.3390/ma17071621
58. Zanesco, A. P., Denkova, E., & Jha, A. P. (2025). Mind-wandering increases in frequency over time during task performance: An individual-participant meta-analytic review.*Psychological Bulletin*, *151*(2), 217–239. https://doi.org/10.1037/bul0000424
59. Zanesco, A. P., Denkova, E., & Jha, A. P. (2025). Mind-wandering increases in frequency over time during task performance: An individual-participant meta-analytic review. *Psychological bulletin*, *151*(2), 217–239. https://doi.org/10.1037/bul0000424
60. Zhoudong, Z., Yiyun, W., Jie, J., Zitong, W., Tao, S., Shufan, R., Na, Y., Sheng, T., Huanqiu, L. (2026) Machine Learning Applied in Small Molecule Drug Discovery: Models, Strategies, and Future Prospects. *Current Topics in Medicinal Chemistry*. https://doi.org/10.2174/0115680266430631251126111506
61. Arayici, Y., Thurairajah, N., & Kumar, B. (2025). Sustainable Communities through Digital Transformation (1st ed.). Routledge. https://doi.org/10.1201/9781003380559
62. Mater, A. C., & Coote, M. L. (2019). Deep Learning in Chemistry. *Journal of chemical information and modeling*, *59*(6), 2545–2559. https://doi.org/10.1021/acs.jcim.9b00266
63. El-Tanani, M., Rabbani, S. A., Wali, A. F., Muhana, F., El-Tanani, Y., & Kumar, R. (2026). Harnessing Machine Learning for Accelerated Drug Discovery: Opportunities and Unmet Challenges. *Pharmaceuticals*, *19*(6), 810. https://doi.org/10.3390/ph19060810
64. Page, M. J., McKenzie, J. E., Bossuyt, P. M., Boutron, I., Hoffmann, T. C., Mulrow, C. D., Shamseer, L., Tetzlaff, J. M., Akl, E. A., Brennan, S. E., Chou, R., Glanville, J., Grimshaw, J. M., Hróbjartsson, A., Lalu, M. M., Li, T., Loder, E. W., Mayo-Wilson, E., McDonald, S., McGuinness, L. A., … Moher, D. (2021). The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. *BMJ (Clinical research ed.)*, *372*, n71. https://doi.org/10.1136/bmj.n71
65. Manan, A., Baek, E., Ilyas, S., & Lee, D. (2025). Digital Alchemy: The Rise of

Machine and Deep Learning in Small-Molecule Drug Discovery. *International Journal of Molecular Sciences*, *26*(14), 6807. https://doi.org/10.3390/ijms26146807

66. Stokes, J. M., Yang, K., Swanson, K., Jin, W., Cubillos-Ruiz, A., Donghia, N. M., ... & Collins, J. J. (2020). A deep learning approach to antibiotic discovery. *Cell*, *180*(4), 688-702. https://doi.org/10.1016/j.cell.2020.01.021
67. Rajpoot K. (2025). Role of Artificial Intelligence in Nanomedicine and Organ-specific Therapy: An Updated Review. *Current drug targets*, *26*(13), 921–953. https://doi.org/10.2174/0113894501394785250715165404
68. Abbas, K., Hao, C., Yong, X., Hasan, M. K., Islam, S., & Rahman, A. H. A. (2025). Graph neural network-based drug-drug interaction prediction. *Scientific reports*, *15*(1), 30340. https://doi.org/10.1038/s41598-025-12936-1
69. Tong, X., Liu, X., Tan, X., Li, X., Jiang, J., Xiong, Z., Xu, T., Jiang, H., Qiao, N., & Zheng, M. (2021). Generative Models for De Novo Drug Design. *Journal of Medicinal Chemistry*, *64*(19), 14011–14027. https://doi.org/10.1021/acs.jmedchem.1c00927
70. Jiménez-Luna, J., Grisoni, F., & Schneider, G. (2020). Drug discovery with explainable artificial intelligence. *Nature Machine Intelligence*, *2*(10), 573-584. https://doi.org/10.1038/s42256-020-00236-4
71. Venkataraman, M., Rao, G. C., Madavareddi, J. K., & Maddi, S. R. (2025). Leveraging machine learning models in evaluating ADMET properties for drug discovery and development. *ADMET & DMPK*, *13*(3), 2772. https://doi.org/10.5599/admet.2772
72. Tan, X., Su, A. T., Hajiabadi, H., Tran, M., & Nguyen, Q. (2021). Applying Machine Learning for Integration of Multi-Modal Genomics Data and Imaging Data to Quantify Heterogeneity in Tumour Tissues. *Methods in molecular biology (Clifton, N.J.)*, *2190*, 209–228. https://doi.org/10.1007/978-1-0716-0826-5_10
73. Corti, C., Cobanaj, M., Dee, E. C., Criscitiello, C., Tolaney, S. M., Celi, L. A., & Curigliano, G. (2023). Artificial intelligence in cancer research and precision medicine: Applications, limitations and priorities to drive transformation in the delivery of equitable and unbiased care. *Cancer treatment reviews*, *112*, 102498. https://doi.org/10.1016/j.ctrv.2022.102498
74. Ching, T., Himmelstein, D. S., Beaulieu-Jones, B. K., Kalinin, A. A., Do, B. T., Way, G. P., Ferrero, E., Agapow, P. M., Zietz, M., Hoffman, M. M., Xie, W., Rosen, G. L., Lengerich, B. J., Israeli, J., Lanchantin, J., Woloszynek, S., Carpenter, A. E., Shrikumar, A., Xu, J., Cofer, E. M., … Greene, C. S. (2018). Opportunities and obstacles for deep learning in biology and medicine. *Journal of the Royal Society, Interface*, *15*(141), 20170387. https://doi.org/10.1098/rsif.2017.0387
75. Qiao, Z., Nie, W., Vahdat, A., Nie, W., Vahdat, A., Miller III, T. F. & Anandkumar A. (2024). State-specific protein–ligand complex structure prediction with a multiscale deep generative model. *Nat Mach Intell* 6, 195–208. https://doi.org/10.1038/s42256-024-00792-z
76. Char, D. S., Shah, N. H., & Magnus, D. (2018). Implementing Machine Learning in Health Care - Addressing Ethical Challenges. *The New England journal of medicine*, *378*(11), 981–983. https://doi.org/10.1056/NEJMp1714229
77. Zhang, H., Ahearn, T. U., Lecarpentier, J., Barnes, D., Beesley, J., Qi, G., Jiang, X., O'Mara, T. A., Zhao, N., Bolla, M. K., Dunning, A. M., Dennis, J., Wang, Q., Ful, Z. A., Aittomäki, K., Andrulis, I. L., Anton-Culver, H., Arndt, V., Aronson, K. J., Arun, B. K., … García-Closas, M. (2020). Genome-wide association study identifies 32 novel breast cancer susceptibility loci from overall and subtype-specific analyses. *Nature genetics*, *52*(6), 572–581. https://doi.org/10.1038/s41588-020-0609-2
78. Kennedy, S. M., K, A., K, P., & Rb, J. R. (2025). Artificial intelligence and machine learning-driven design of self-healing biomedical composites. *Expert review of medical devices*, *22*(8), 787–805. https://doi.org/10.1080/17434440.2025.2520291
79. Gianfrancesco, M. A., Tamang, S., Yazdany, J., & Schmajuk, G. (2018). Potential Biases in Machine Learning Algorithms Using Electronic Health Record Data. *JAMA internal medicine*, *178*(11), 1544–1547. https://doi.org/10.1001/jamainternmed.2018.3763
80. Gee, G. C., & Payne-Sturges, D. C. (2004). Environmental health disparities: a framework integrating psychosocial and environmental concepts. *Environmental health perspectives*, *112*(17), 1645–1653. https://doi.org/10.1289/ehp.7074
81. Chen, J., Atkinson, R. W., Andersen, Z. J., Oftedal, B., Stafoggia, M., Lim, Y. H., Bekkevold, T., Krog, N. H., Renzi, M., Zhang, J., Bauwelinck, M., Janssen, N., Strak, M., Forastiere, F., de Hoogh, K.,

Rodopoulou, S., Katsouyanni, K., Raaschou-Nielsen, O., Samoli, E., Brunekreef, B., … Vienneau, D. (2024). Long-term exposure to ambient air pollution and risk of lung cancer - A comparative analysis of incidence and mortality in four administrative cohorts in the ELAPSE study. *Environmental research*, *263*(Pt 3), 120236. https://doi.org/10.1016/j.envres.2024.120236

82. Ingelman-Sundberg M. (2002). Polymorphism of cytochrome P450 and xenobiotic toxicity. *Toxicology*, *181-182*, 447–452. https://doi.org/10.1016/s0300-483x(02)00492-4
83. Mehrabi, N., Morstatter, F., Saxena, N., Lerman, K., Galstyan, A. (2022). A survey on bias and fairness in machine learning algorithms across medical sub-domains. *ACM Computing Surveys*, *54*(6), 1–35. https://doi.org/10.1145/3457607
84. Bellamy, R. K., et al. (2023). AI Fairness 360: An extensible toolkit for measuring and mitigating algorithmic bias in healthcare datasets. *IBM Journal of Research and Development*, *67*(2), 6:1-6:14. https://doi.org/10.1147/JRD.2019.2942287
85. Zafar, M. B., Valera, I, Rogriguez, M. G., Gummadi, K. P. (2024). Fairness constraints: A mechanism for fair classification and demographic parity in clinical neural architectures. *Journal of Machine Learning Research*, *25*(42), 1–48. https://proceedings.mlr.press/v54/zafar17a.html
86. Yadav, S. K., Shahi, S. (2024). Safe Human-Robot Collaboration in Dynamic Environments: An AI-Powered Situation Awareness Perspective. International Journal of Tropical Medicine, 19 (4), 92-98, https://doi.org/10.36478/makijtm.2024.4.92.98
87. Singh, S. K., Shahi, S. (2024). AI for Earth: Enhancing Environmental Sustainability with Technology. Research Journal of Medical Sciences, 18 (11), 219-225. https://doi.org/10.36478/makrjms.2024.11.219.225
88. Shahi, S., Singh, S. K. (2024). Microfluidic Devices with Nano-sensors for Infectious Disease Detection. Nanotechnology Perceptions, 20 (14s), 3369-3377. https://nano-ntp.com/index.php/nano/article/view/3295/2466
89. Rajkomar, A., Hardt, M., Howell, M. D., Corrado, G., & Chin, M. H. (2018). Ensuring Fairness in Machine Learning to Advance Health Equity. *Annals of internal medicine*, *169*(12), 866–872. https://doi.org/10.7326/M18-1990
90. US Food and Drug Administration. (2024). *Artificial Intelligence and Medical Products: Maintaining Equity, Safety, and Quality Lifecycle Governance*. FDA Center for Drug Evaluation and Research White Paper. https://www.fda.gov/media/177030/download
91. European Medicines Agency. (2024). *Reflection Paper on the Use of Artificial Intelligence and Machine Learning in the Lifecycle of Medicinal Products*. EMA/CHMP/CVMP/83833/2023. https://l1nq.com/cxcgtp2
92. PDQ Pediatric Treatment Editorial Board. (2025). Childhood Acute Lymphoblastic Leukemia Treatment (PDQ®): Health Professional Version. In *PDQ Cancer Information Summaries*. National Cancer Institute (US). https://pubmed.ncbi.nlm.nih.gov/26389206/
93. Price, W. N., 2nd, & Cohen, I. G. (2019). Privacy in the age of medical big data. *Nature medicine*, *25*(1), 37–43. https://doi.org/10.1038/s41591-018-0272-7
94. Kaissis, G. A., Makowski, M. R., Ruckert, D., Braren, R. F. (2020). Secure, privacy-preserving and federated machine learning in medical imaging and oncology analytics. *Nature Machine Intelligence*, *2*(6), 305–311.https://doi.org/10.1038/s42256-020-0186-1
95. Subbaswamy, A., & Saria, S. (2020). From development to deployment: dataset shift, causality, and shift-stable models in health AI. *Biostatistics (Oxford, England)*, *21*(2), 345–352. https://doi.org/10.1093/biostatistics/kxz041
96. Han, X., Aslanian, A., & Yates, J. R., 3rd (2008). Mass spectrometry for proteomics. *Current opinion in chemical biology*, *12*(5), 483–490. https://doi.org/10.1016/j.cbpa.2008.07.024
97. Barisoni, L., Lafata, K. J., Hewitt, S. M., Madabhushi, A., & Balis, U. G. J. (2020). Digital pathology and computational image analysis in nephropathology. *Nature reviews. Nephrology*, *16*(11), 669–685. https://doi.org/10.1038/s41581-020-0321-6
98. Huang, Q., Li, Y., Huang, Y., Wu, J., Bao, W., Xue, C., Li, X., Dong, S., Dong, Z., & Hu, S. (2025). Advances in molecular pathology and therapy of non-small cell lung cancer. *Signal transduction and targeted therapy*, *10*(1), 186. https://doi.org/10.1038/s41392-025-02243-6
99. Arya, K. Y., Tripathy, P., Awasthi, S., Shahi, S. (2024). A Comparative Study of

Traditional and Digital Approaches to Chemistry Education. African Journal of Biomedical Research. 27 (4s), 9874-9879, https://doi.org/10.53555/AJBR.v27i4S.5561

100. Geng, J., Sui, X., Du, R., Feng, J., Wang, R., Wang, M., Yao, K., Chen, Q., Bai, L., Wang, S., Li, Y., Wu, H., Hu, X., & Du, Y. (2024). Localized fine-tuning and clinical evaluation of deep-learning based auto-segmentation (DLAS) model for clinical target volume (CTV) and organs-at-risk (OAR) in rectal cancer radiotherapy. *Radiation oncology (London, England)*, *19*(1), 87. https://doi.org/10.1186/s13014-024-02463-0
101. Anand, S., Choubey, S., Sharma, V., Shahi, S. (2024). From Tradition to Modern Medicine: The Evolving Role of Medicinal Plants. Cuestiones De Fisioterapia, 53 (3), 545-554. https://doi.org/10.48047/fwgmdt56
102. Sheller, M. J., Edwards, B., Reina, G. A., Martin, J., Pati, S., Kotrotsou, A., Milchenko, M., Xu, W., Marcus, D., Colen, R. R., & Bakas, S. (2020). Federated learning in medicine: facilitating multi-institutional collaborations without sharing patient data. *Scientific reports*, *10*(1), 12598. https://doi.org/10.1038/s41598-020-69250-1
103. Dayan, I., Roth, H. R., Zhong, A., Harouni, A., Gentili, A., Abidin, A. Z., Liu, A., Costa, A. B., Wood, B. J., Tsai, C. S., Wang, C. H., Hsu, C. N., Lee, C. K., Ruan, P., Xu, D., Wu, D., Huang, E., Kitamura, F. C., Lacey, G., de Antônio Corradi, G. C., … Li, Q. (2021). Federated learning for predicting clinical outcomes in patients with COVID-19. *Nature medicine*, *27*(10), 1735–1743. https://doi.org/10.1038/s41591-021-01506-3
104. Guzmán Gómez, R., Lopez Lopez, G., Alvarado, V. M., Lopez Lopez, F., Esqueda Cisneros, E., & López Moreno, H. (2025). Deep Learning Approaches for Automated Prediction of Treatment Response in Non-Small-Cell Lung Cancer Patients Based on CT and PET Imaging. *Tomography (Ann Arbor, Mich.)*, *11*(7), 78. https://doi.org/10.3390/tomography11070078
105. Pati, S., Baid, U., Edwards, B., Sheller, M., Wang, S. H., Reina, G. A., Foley, P., Gruzdev, A., Karkada, D., Davatzikos, C., Sako, C., Ghodasara, S., Bilello, M., Mohan, S., Vollmuth, P., Brugnara, G., Preetha, C. J., Sahm, F., Maier-Hein, K., Zenk, M., … Bakas, S. (2022). Federated learning enables big data for rare cancer boundary detection. *Nature communications*, *13*(1), 7346. https://doi.org/10.1038/s41467-022-33407-5
106. Shahi, S., Singh,S. K. (2025) Microplastic Pollution: A Growing Threat to Marine Ecosystems. Cuestiones De Fisioterapia, 54 (2), 3077-3089. https://doi.org/10.48047/yrf7m803
107. Karimireddy, S. P., Satyen, K., Mehryar, M., Reddi, S. J. Stich, Sebastian U., Suresh, A. T. (2021). Scaffold: Stochastic controlled averaging for federated learning under severe label distribution shift. *International Conference on Machine Learning (ICML)*, *42*, 5132–5143. https://infoscience.epfl.ch/handle/20.500.14299/181669
108. Shahi, S., Singh, S. K. (2025). Phytochemical Analysis and Characterization of Cancer-Cure Medicinal Plants: A Comprehensive Review. International Journal of Environmental Sciences, 11 (10s), 592-606. https://doi.org/10.64252/rg3fjw45
109. Aburass, S., Dorgham, O., Al Shaqsi, J., Abu Rumman, M., & Al-Kadi, O. (2025). Vision Transformers in Medical Imaging: a Comprehensive Review of Advancements and Applications Across Multiple Diseases. *Journal of imaging informatics in medicine*, *38*(6), 3928–3971. https://doi.org/10.1007/s10278-025-01481-y
110. Obbalareddy, S., Shahi, S., Dubey, A., Sehajpal, S., Samanthula, K. S., Saran, N., Rani, D. (2025). IMPLEMENTING ARTIFICIAL INTELLIGENCE IN DRUG RESEARCH DEVELOPMENT, BULLETIN OF STOMATOLOGY AND MAXILLOFACIAL SURGERY, 21 (6), 361-384, https://doi.org/10.58240/1829006X-2025.21.6-361
111. Gentry, C. (2009). Fully homomorphic encryption using an ideal lattice architecture. *ACM Symposium on Theory of Computing (STOC)*, *41*, 169–178. https://doi.org/10.1145/1536414.1536440
112. Liu, C., Li, K., Sui, Y., Liu, H., Zhang, Y., Lu, Y., Lu, W., Chen, Y., Wang, G., Xu, S., Xiang, T., Cai, Y., & Huang, K. (2023). Different gene alterations in patients with non-small-cell lung cancer between the eastern and southern China. *Heliyon*, *9*(10), e20171. https://doi.org/10.1016/j.heliyon.2023.e20171
113. Guo, Y., Jiang, L., Alkali, M., George, S. H. L., Jones, P. D., Leng, S., & Ye, F. (2025). Aflatoxin Exposure, Chronic Hepatitis Infection, and HLA Diversity Converge to Shape Patient Survival in Hepatocellular Carcinoma. *Cancer epidemiology, biomarkers & prevention: a publication of the American Association for Cancer Research, cosponsored by the American Society of Preventive Oncology*, *34*(11), 1888–1894.

https://doi.org/10.1158/1055-9965.EPI-25-0795

114. Knott, B., et al. (2021). CryptTen: A high-performance, researcher-friendly framework for secure machine learning. *Advances in Information Processing Systems (NeurIPS)*, *34*, 1241–1253. https://proceedings.neurips.cc/paper_files/paper/2021
115. Cho, H., Wu, D. J., & Berger, B. (2018). Secure genome-wide association analysis using multiparty computation. *Nature biotechnology*, *36*(6), 547–551. https://doi.org/10.1038/nbt.4108
116. Froelicher, D., Troncoso-Pastoriza, J. R., Raisaro, J. L., Cuendet, M. A., Sousa, J. S., Cho, H., Berger, B., Fellay, J., & Hubaux, J. P. (2021). Truly privacy-preserving federated analytics for precision medicine with multiparty homomorphic encryption. *Nature communications*, *12*(1), 5910. https://doi.org/10.1038/s41467-021-25972-y
117. Singh, S. K., Tripathy, P., Singh, A., Shahi, S. (2026). Artificial Intelligence-Integrated Ethnomedicine Of Odisha: Chemotypes, Bioactive Compounds, And Translational Drug Discovery From Tribal Knowledge Systems. SCIENTIFIC CULTURE, 12 (4), 5846-5862.https://sci-cult.net/index.php/cult/article/view/4873/2855
118. Obbalareddy, S., Dubey, A., Kumari, M., Mukherjee, C., Shahi, S., Dutta, S. (2026) A Systematic Review of Vaccination, Prevention, and Public Health Measures for the Human Papillomavirus Epidemic Among Patients in India. International Journal of Pharmaceutical Investigation, 16 (1), 58-65, https://doi.org/10.5530/ijpi.20260462
119. Watson, J., et al. (2024). Hardware acceleration metrics for lattice-based homomorphic encryption schemes inside cloud infrastructures. *IEEE Transactions on Computers*, *73*(5), 1340–1354.
120. Watson, J. L., Wagh, S., & Popa, R. A. (2022). Piranha: A {GPU} platform for secure computation. In *31st USENIX Security Symposium (USENIX Security 22)* (pp. 827-844).
121. Al Badawi, A., & Faizal Bin Yusof, M. (2024). Private pathological assessment via machine learning and homomorphic encryption. *BioData mining*, *17*(1), 33. https://doi.org/10.1186/s13040-024-00379-9
122. Zsidó, B. Z., Bayarsaikhan, B., Börzsei, R., Szél, V., Mohos, V., & Hetényi, C. (2023). The Advances and Limitations of the Determination and Applications of Water Structure in Molecular Engineering. *International Journal of Molecular Sciences*, *24*(14), 11784. https://doi.org/10.3390/ijms241411784
123. Bender, A., & Cortés-Ciriano, I. (2021). Artificial intelligence in drug discovery: what is realistic, what are illusions? Part 1: Ways to make an impact, and why we are not there yet. *Drug discovery today*, *26*(2), 511–524. https://doi.org/10.1016/j.drudis.2020.12.009
124. Sala, D., Batebi, H., Ledwitch, K., Hildebrand, P. W., & Meiler, J. (2023). Targeting in silico GPCR conformations with ultra-large library screening for hit discovery. *Trends in pharmacological sciences*, *44*(3), 150–161. https://doi.org/10.1016/j.tips.2022.12.006
125. Hou, Z., Li, Y., Zhai, H., Luo, J., Ding, Y., & Pan, Y. (2025). MEGDTA: multi-modal drug-target affinity prediction based on protein three-dimensional structure and ensemble graph neural network. *BMC genomics*, *26*(1), 738. https://doi.org/10.1186/s12864-025-11943-w
126. You, Y., Lai, X., Pan, Y., Zheng, H., Vera, J., Liu, S., Deng, S., & Zhang, L. (2022). Artificial intelligence in cancer target identification and drug discovery. *Signal transduction and targeted therapy*, *7*(1), 156. https://doi.org/10.1038/s41392-022-00994-0

How to cite this article: Sanyogita Shahi, Shirish Kumar Singh, Vinod Kumar Choudhary, AI-DRIVEN PARADIGMS IN PRECISION ONCOLOGY: MAPPING THE CONVERGENCE OF MULTI-OMICS DRUG DISCOVERY, MEDICINAL CHEMISTRY, AND ENVIRONMENTAL CARCINOGENESIS, Asian J. Med. Res. Health Sci., 2026; 4 (2):2180-2202.
Source of Support: Nil, Conflicts of Interest: None declared.