

Recent Advances and Emerging Directions in Machine Learning-Based Breast Cancer Drug Discovery: A Comprehensive Review

Yan Cheng, Jilin Kong, Xiangjuan Liu, Shuyan Li

Breast Surgery Department of Yantai Hospital, Yantai, Shandong, 264100, People's Republic of China

Correspondence: Shuyan Li, Breast Surgery Department of Yantai Hospital, Yantai, Shandong, 264100, People's Republic of China, Email fv8450@163.com

Abstract: Machine learning (ML) has emerged as a powerful technique for multiple stages of breast cancer drug discovery, from target identification to compound prioritization and patient stratification. This article presents a narrative review of recent advances in ML-driven strategies applied to breast cancer drug discovery, with a focus on methods, data resources, and translational relevance. We systematically synthesize representative studies employing supervised and unsupervised learning, deep neural networks, generative models, and multi-omics integration to address key challenges in breast cancer therapeutics. Particular attention is given to ML approaches for biomarker discovery, drug–target interaction prediction, molecular design, and drug response modeling across breast cancer subtypes. The review also summarizes widely used public datasets, including genomic, transcriptomic, pharmacological, and chemical repositories that underpin these approaches. In addition, we discuss reported translational applications, emerging industrial efforts, and critical limitations related to data bias, model generalizability, and clinical applicability. Finally, we outline future directions for improving the robustness, interpretability, and clinical integration of ML-based drug discovery frameworks, aiming to bridge the gap between computational prediction and applicable breast cancer therapies.

Keywords: drug discovery, drug candidates, machine learning, breast cancer, oncology

Introduction

Cancer remains one of the leading causes of death worldwide, with considerable mortality.^{1–4} Among various cancer types, breast cancer is one of the most common malignancies and the leading cause of cancer-related mortality in women worldwide.⁵ The global incidence of breast cancer is rising as of 2022, almost 2.3 million women are diagnosed, and approximately 670,000 die due to breast cancer.^{6,7} Therefore, it is considered among the most frequently diagnosed cancers in 157 of 185 countries.⁸ However, with advancements in therapeutic approaches, the survival rates of cancer patients are enhanced, but cancer cases are also expected to increase.^{6,9} Long term data confirm that the incident cases nearly doubled from 0.88 million in 1990 to over 2.1 million in 2021.¹⁰ The burden varies based on geographical and socioeconomic conditions, and a higher incidence and mortality have been observed in affluent regions.^{8,10} For example, the lifetime breast cancer risk in very high Human Development Index (HDI) countries is approximately 1 in 12, whereas it is approximately 1 in 27 in low HDI countries.⁸ Breast cancer is a heterogeneous and usually classified according to hormone receptor (ie, estrogen/progesterone) and human epidermal growth factor receptor 2 (HER2) status (subtypes: luminal A, luminal B, HER2+, and triple-negative).^{11–13} Each subtype has a specific pathophysiology, gene expression profile, prognostic outcome, and therapeutic responses.¹¹ Furthermore, extensive inter- and intra-tumor heterogeneity makes it more complex to understand the disease etiology and designing effective targeted therapies.¹¹

Despite significant advances in cancer drug discovery and drug delivery technologies, multiple biological, technical, and translational limitations continue to hinder clinical outcomes.^{14–18} Bringing a candidate therapy from bench to bedside typically spans 10–15 years, and oncology has an extraordinarily high attrition rate (often cited as >95%



candidate failures).¹⁹ Early in the development process, selecting and validating targets was difficult. Translation of complex tumor biology into effective therapies was a great challenge requiring strong disease linkage and rigorous preclinical validation.²⁰ In practice, many targets are pursued with only modest evidence of causality, and breast tumor heterogeneity and paucity of predictive biomarkers (aside from HER2) make rational target selection difficult.²¹ Consequently, most preclinical studies have not identified viable drugs. For example, 16 distinct investigational agents targeting the insulin-like growth factor-1 receptor (IGF-1R) pathway have been tested, but none have shown sufficient activity for approval.¹⁹ These problems continue to arise during clinical development. Even agents that reach the trial stage often fail in terms of efficacy or safety endpoints. In a Phase II study, adding the IGF-1R antibody ganitumab to endocrine therapy for metastatic Estrogen Receptor-Positive (ER+) breast cancer did not improve progression-free survival and even worsened overall survival with frequent grade ≥ 3 toxicities.²¹ Likewise, the pan Phosphatidylinositol 3 kinase (pan-PI3K) inhibitor buparlisib gave only a modest progression-free survival (PFS) gain (median 6.9 vs 5.0 months; HR = 0.78),²² but at the cost of significant toxicity, for example, grade 3 ALT (liver enzyme) elevations occurred in 25% of buparlisib patients versus 1% of controls.²² In both cases, the limited benefits failed to justify the adverse effects, and further developments were abandoned. These limitations, in terms of efficiency and safety concerns, are common in oncology trials. The combination of prolonged, expensive trials and frequent late-stage failures indicates that traditional breast cancer studies produce relatively few new drugs despite huge investments and efforts.

In modern drug discovery, artificial intelligence (AI) and machine learning (ML) have increasingly been used to overcome traditional drawbacks. In breast cancer research, deep neural networks are used to model complex structure-activity relationships, which provide accurate predictions of compound efficacy and safety.²³ Currently, pharmaceutical and biotech organizations use ML to expedite virtual compound screening and predict drug-target interactions from vast chemical libraries.^{24,25} In particular, generative AI models (including variational autoencoders and GANs) can propose entirely new small molecules with the desired properties. Moreover, the repetition of reinforcement learning techniques optimizes these candidates for target binding and pharmacokinetic goals.²⁴ These approaches significantly improve hit-to-lead timelines and reduce the dependency on exhaustive laboratory screening systems.²⁴ Similarly, AI approaches are being used to personalize therapy. ML algorithms proficient in tumor genomics, imaging, and clinical data can be used to classify patients with breast cancer into subgroups for treatments. For example, by combining digital pathology and radiomic data, AI models have significantly improved the prediction of PD-L1 biomarker status.^{26–29} Analogous ML systems are under development to match breast cancer subtype profiles with those of targeted therapies. Recent studies have highlighted that AI-driven target prediction and targeted drug generation are among the most promising applications in precision oncology.²⁶ Several AI-driven platforms have been applied to target identification in oncology. For example, in silico Medicine's AI platform was used to analyze large-scale transcriptomic and pathway-level data with the goal of identifying immune related therapeutic targets. It prioritized previously underexplored targets associated with immune signaling pathways, some of which were subsequently evaluated in collaboration with GlaxoSmithKline in preclinical oncology research.²⁴ Similarly, several biotechnology companies, including Exscientia and BenevolentAI, have developed AI based platforms for oncology drug discovery. These platforms apply ML models to integrate chemical, biological, and clinical data for tasks such as target identification, compound prioritization, and preclinical candidate selection in cancer research. While these approaches demonstrated improved efficiency in in silico screening and hypothesis generation. Their direct impact on reducing the development timeline for breast cancer specific therapeutics remains largely supported by preclinical studies and early translational evidence rather than by the completed clinical drug development process.^{23,24}

This study highlights a comprehensive overview of the importance of ML in the drug discovery process for breast cancer. It covers each stage, from the identification of the target and hit discovery through lead optimization to preclinical testing. This highlights recent algorithmic advances, including supervised and unsupervised learning, deep neural architectures, and generative and reinforcement learning models. Moreover, it covers the current results and translational case studies and provides practical insights. It also addresses the main challenges, limitations, and future directions for accelerating AI-driven breast cancer drug discovery.

Machine Learning in the Drug Discovery Pipeline

Traditionally, drug discovery has proceeded through sequential stages, including target identification (finding a disease-related biomolecule), hit discovery (screening for initial drugs), lead optimization (refining drugs for potency and selectivity), and preclinical testing (*in vitro* and animal studies).^{30,31} This process is extremely long and expensive, and often takes more than a decade and billions of dollars.³¹ ML tools are now integrated at each stage, such as ML algorithms that mine complex omics data to nominate drug targets, virtually screen millions of compounds to find hits, and guide chemical modifications to optimize leads. It also predicts the absorption, distribution, metabolism, excretion, and toxicity (ADMET) of drugs before testing.^{31,32} These sequential ML applications have the potential to streamline and to accelerate the overall drug discovery process.^{31,32}

Target Identification

ML has become pivotal in the identification of targets, especially for complex diseases, such as breast cancer. By sifting through high-dimensional genomic and proteomic data, ML algorithms can identify genes that are critical to disease survival. Feature selection methods, such as support vector machines (SVM) with recursive feature elimination or random forests with importance ranking, help to highlight candidate targets by identifying genomic/proteomic features that distinguish cancer from normal or resistant phenotypes. For instance, an integrative ML process that combined ensemble learning and metaheuristic optimization identified 35 key genes as robust biomarkers and therapeutic targets in breast cancer.³³ Notably, this data-driven approach rediscovered well-known targets like TOP2A (topoisomerase II α), which is the focus of several breast cancer drugs.^{33,34} The fact that many of these ML-prioritized genes are druggable underscores how ML can identify clinically relevant targets.³³

Network biology offers a powerful avenue for ML-driven target discovery. Cancer arises from perturbed interaction networks, and ML can exploit the network topology to identify influential “hub” genes/proteins.³⁵ Studies have applied classifiers to protein-protein interaction networks to identify central nodes linked to disease.^{35–38} For example, an adaptive decision tree model using breast cancer network data found that connectivity features (such as node degree and neighbor metrics) best separated cancer-associated proteins from others,³⁹ effectively flagging hub proteins as potential targets. Similarly, network algorithms augmented by ML (eg, decision trees with particle swarm optimization) have been used to prioritize cancer treatment targets by avoiding local optima during feature selection.^{40,41} Deep learning further enhances target identification by integrating heterogeneous data sources.^{42,43} Graph neural networks and other deep models can learn complex patterns across multi-omics and predict novel gene-disease associations.³⁵ In practice, AI-driven platforms such as Open Targets use approaches such as gradient-boosted trees (XGBoost) to combine genetic associations and functional data, and successfully score candidate genes at disease loci.⁴⁴ In particular, genomics-guided ML is promising, as approximately two-thirds of recently approved drug target genes are supported by human genetic evidence.⁴⁴ However, proteomics-based ML analyses reveal therapeutic vulnerabilities from a protein perspective; for example, a phosphoproteomic study identified focal adhesion kinase 2 abbreviated as FAK2, the main driver of tamoxifen-resistant breast cancer. Experimental validation confirmed that FAK2 is a viable therapeutic target for hormone-refractory tumors.⁴⁵

The combination of ML approaches, from traditional classifiers (SVM and random forest) to deep learning, and network-informed modeling enables target identification. These methods systematically filter the noise in big omics data to identify new and previously overlooked molecular targets. By employing patterns in gene expression, mutations, protein networks, and patient data, ML accelerates the discovery and prioritization of drug targets in breast cancer. This provides a stronger foundation for subsequent drug development steps.^{33,35}

Hit Discovery

ML transforms hit discovery through efficient virtual screening of vast chemical libraries. Deep-learning-based scoring strategies (deep docking) can accelerate structure-based virtual screening 100-fold and enrich true hits by several orders of magnitude. This allows billion-scale libraries to be screened without prohibitive computational cost.⁴⁶ Deep neural networks and other ML models have been successfully applied to identify hits against breast cancer targets, often

surpassing classical docking-based scoring in prioritizing true active compounds.⁴⁷ For instance, Shafiq et al integrated a deep-learning classifier with molecular docking to discover isoform-selective PI3K α inhibitors.⁴⁸ Their pipeline filtered a large library of 662 candidate inhibitors, and subsequent docking yielded 12 high-affinity hits, including one compound previously confirmed to be active *in vitro* and several with stronger predicted binding than the approved PI3K α inhibitor, alpelisib.⁴⁸ Similarly, Jaiswal et al employed a convolutional neural network (CNN) to virtually screen PI3K α inhibitors in triple-negative breast cancer.⁴⁹ CNN-based screening of a chemical library identified 12 candidate hits, four of which were validated as potent PI3K α inhibitors *in vitro*.⁴⁹ Graph neural networks (GNNs) have likewise shown promise in hit identification: Mahanta et al developed graph-based models for TNBC that achieved area under the curve (AUC) values up to 0.82 and identified key structural fragments associated with anti-cancer activity, with the top-performing message-passing neural network predicting active compounds with up to 97% accuracy on an external test set.⁵⁰ A schematic representation of ML-based hit discovery is shown in Figure 1.

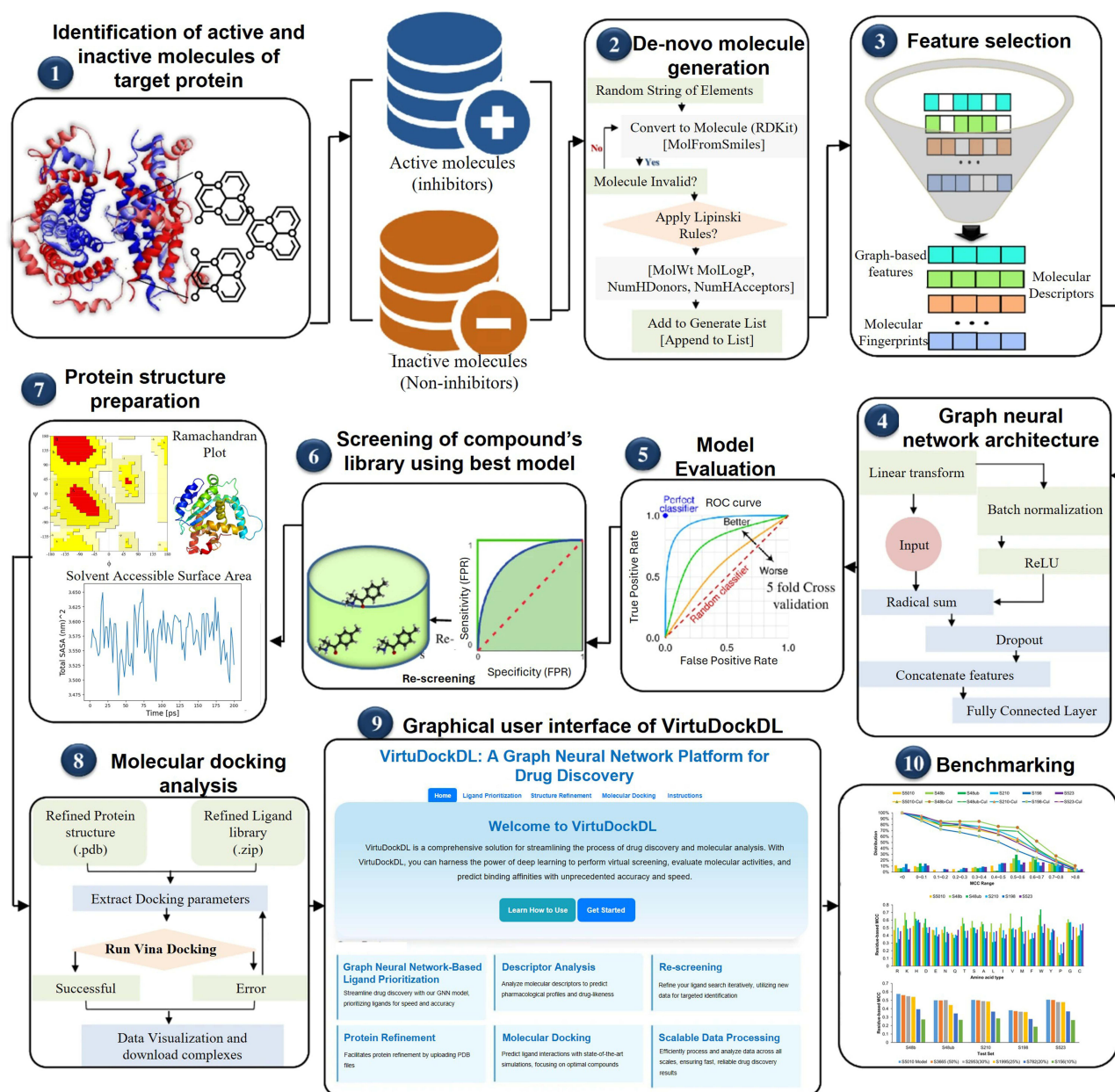


Figure 1 A schematic representation of ML-based hit discovery.⁵¹

Lead Optimization

In the lead optimization stage of breast cancer drug discovery, machine learning (ML) techniques are pivotal for refining candidate compounds to maximize efficacy and drug-likeness. A critical focus is the early prediction of ADMET properties, as undesirable ADMET profiles are the leading cause of late-stage failure.^{52,53} ML models (eg, random forests, support vector machines, and neural networks) now enable *in silico* ADMET screening of lead libraries with high throughput.^{54–57} This approach has limitations, such as poor solubility or toxicity before clinical trials, which saves time and resources. Advanced deep learning architectures (ie, GNNs) further improve ADMET and pharmacokinetic predictions by directly learning from molecular structures.⁵² Similarly, ML-driven structure activity relationship (SAR) modeling guides the chemical optimization of the leads. Data-driven quantitative structure-activity relationships, abbreviated as QSAR models, can predict a compound's bioactivity from its structure, which helps chemists prioritize modifications that enhance efficiency and selectivity. For example, a recent ML-based QSAR study on a flavone derivative library achieved high predictive power ($R^2 = 0.82$) for cytotoxicity against MCF-7 breast cancer cells. This has enabled the design of new analogs with improved anti-breast cancer activity and minor toxicity to normal cells.⁵⁸ Moreover, transfer learning techniques allow leveraging large chemical and bioactivity datasets to boost the modeling performance for novel targets,⁵⁹ improving SAR predictions, even when data on a specific breast cancer target are limited. Beyond prediction, generative models are used for *de novo* compound design, and deep generative neural networks (variational autoencoders, GANs) combined with reinforcement learning (RL) can autonomously propose new molecules optimized for desired properties.⁶⁰ Notably, the generative model GENTRL identified potent DDR1 kinase inhibitors within a few weeks,⁶⁰ and the MORLD platform demonstrated that an RL-based agent can optimize lead molecules (eg, refining a kinase inhibitor's structure for better binding and drug-like properties) within days using only the 3D structure of the target.⁶⁰ Such ML-driven lead optimization strategies significantly accelerate the development of promising breast cancer therapeutics. This was achieved by balancing anti-tumor potency with pharmacological and safety profiles to improve the odds of clinical success.

Preclinical Testing

During the preclinical testing of breast cancer drug candidates, ML provides critical *in silico* predictions to guide safety and efficacy. ML models can predict diverse toxic endpoints (eg, acute systemic toxicity, cardiotoxicity, and hepatotoxicity)⁶¹ and flagging compounds, which may cause adverse side effects prior to *in vivo* studies. This accelerates the risk by reducing the reliance on trial-and-error animal testing. ML also aids in modeling drug-target interactions and pharmacokinetics, such as deep learning, and ensemble algorithms predict binding affinities and off-target profiles of lead compounds, as well as absorption, distribution, metabolism, and excretion properties.²⁴ Different techniques from tree-based models to deep neural networks and even transfer learning to leverage large datasets have achieved strong performance in bioactivity and ADMET prediction.⁶² This allows for optimization of breast cancer treatment, leading to potency and tolerability. Moreover, ML enables the analysis of high-throughput preclinical data, such as deep learning, to analyze histopathological images from animal studies to automatically detect and score tissue lesions,⁶³ which is consistent with expert pathologists. Generative models can simulate complex experimental readouts, such as GAN-based Tox-GAN systems, which can generate synthetic *in vivo*-like transcriptomic profiles under various dosing conditions with high fidelity to real gene expression data.⁶⁴ Integrating multi modal data (omics and imaging) assists in elucidating drug mechanisms and toxicity biomarkers in breast cancer models. ML supports dosing regimen design and response prediction in preclinical settings. For instance, a hybrid mechanistic ML study in a breast cancer mouse model with varied neoadjuvant dosing indicated that treatment significantly shrank primary tumors but failed to prevent postsurgical metastatic progression.⁶⁵ Emerging reinforcement learning approaches are being explored to optimize adaptive dosing schedules *in silico*.⁶⁶ These capabilities demonstrate how incorporating ML into preclinical testing accelerates the prioritization of breast cancer therapeutics for clinical development.

Recent Advances in ML Algorithms

ML methods have been rapidly adopted in breast cancer drug discovery. Modern approaches combine traditional ML with novel algorithms to analyze high-dimensional omics data. For instance, hybrid metaheuristic feature-selection techniques have been applied to transcriptomic profiles to identify druggable genes.³³ Simultaneously, the ML-driven

integration of radiomics and genomics enables non-invasive subtyping and therapy design.^{67,68} Overall, recent studies have reported that multi-modal ML pipelines can predict patient-specific drug responses and prioritize therapeutic targets with high accuracy.^{68,69} These approaches often employ advanced supervised and unsupervised algorithms and specialized feature engineering to cope with the complexity and heterogeneity of breast tumors. The following subsections provide detailed examples of these methods, including algorithmic advancements tailored to breast cancer data.

Supervised and Unsupervised Learning Approaches

Supervised learning is widely used to predict drug activity and responses. Random forests, support vector machines, and gradient-boosting models have been trained on breast cancer omics and chemical data to associate gene features with drug sensitivity.^{33,69} For example, an XGBoost-style ensemble was recently shown to predict drug responses by weighing the gene-expression and drug descriptors. Similarly, specialized feature-selection algorithms enhance performance. Ahmed et al applied binary grey wolf optimization with simulated annealing to select predictive gene signatures for drug target discovery, achieving near-perfect cross-validated accuracy.³³ In practice, supervised pipelines often incorporate omics feature engineering to reduce noise and identify predictive biomarkers. The extracted features include differentially expressed genes (DEGs), pathway activity scores, and network-central proteins that were correlated with drug outcomes.

Unsupervised learning plays an important role in the identification of structures without labels. Clustering algorithms (k-means and hierarchical clustering) applied to tumor gene-expression or proteomic data can reveal novel subtypes and identify candidate drug targets that are specific to each cluster. For instance, unsupervised analysis of combined gene-expression datasets often group samples into clinically relevant subtypes (eg, luminal vs basal-like), which can guide subtype-specific therapy choices.^{33,69} Advanced unsupervised methods (eg, non-negative matrix factorization, spectral clustering, and graph-based community detection) can further extract latent features (metagenes) from multi-omics profiles, highlighting co-regulated genes or metabolomic pathways that may serve as drug targets. A recent study by Malik et al fused two transcriptomic studies and applied consensus feature selection, after filtering DEGs, they performed hierarchical clustering of expression patterns to stratify cases and identified 35 robust genes as biomarkers.⁶⁹ Such unsupervised pipelines often precede supervised modeling, providing hypotheses and features for downstream classification (eg, clustering responders versus non-responders by phenotype). Overall, modern pipelines iteratively combine clustering and classification; clustering highlights subtype-specific biology and isolates noisy samples, followed by supervised learning (eg, random forest or lightGBM) to build predictive models of curated features.^{33,69} Together, these algorithms accelerate target identification from high-dimensional data, while mitigating noise and overfitting. [Figure 2](#) presents some commonly utilized examples of supervised and unsupervised ML approaches for drug discovery.⁷⁰

Deep Learning and Neural Architectures

Deep learning architectures are increasingly used to model complex data in breast cancer discovery.⁷¹ Convolutional neural network (CNNs) processes imaging and molecular data. For example, CNNs analyze high-throughput microscopy images of breast cancer cells or screen chemical compound images to predict viability and phenotypes.^{72,73} CNN-based QSAR models use molecular fingerprints or 3D structural grids to score ligand-target binding.⁷⁴ A recent explainable architecture combined a GNN of drug molecules with a CNN for tumor gene-expression. In the explainable graph-based drug response prediction (XGDP) model, a GNN learns the latent features from the molecular graph of each drug, whereas a CNN compresses the transcriptomic profile of a cancer cell line. These feature maps are fused via multi-head attention to predict drug responses.⁷⁵ This architecture improves upon earlier models by jointly interpreting the molecular structure and cellular context in an end-to-end fashion. Other neural models include graph-based attention models, which encode 3D pharmacophores into topology-preserving vectors, and variational autoencoders (VAEs), which compress chemical structures or biological networks into latent spaces. For instance, GraphDRP replaced CNNs with GNNs to better capture drug structures, achieving higher accuracy in sensitivity prediction.⁷⁵ Self-attention mechanisms (eg, SWNet) incorporate drug similarity graphs when encoding cell-line features.⁷⁶ Natural language transformers applied to SMILES strings or protein sequences have also been used to predict the ADMET properties and target interactions. In silico tools, such as DeepChem, implement many architectures for chemoinformatics.^{77,78} Recent advances have also focused on explainability: integrated gradient and

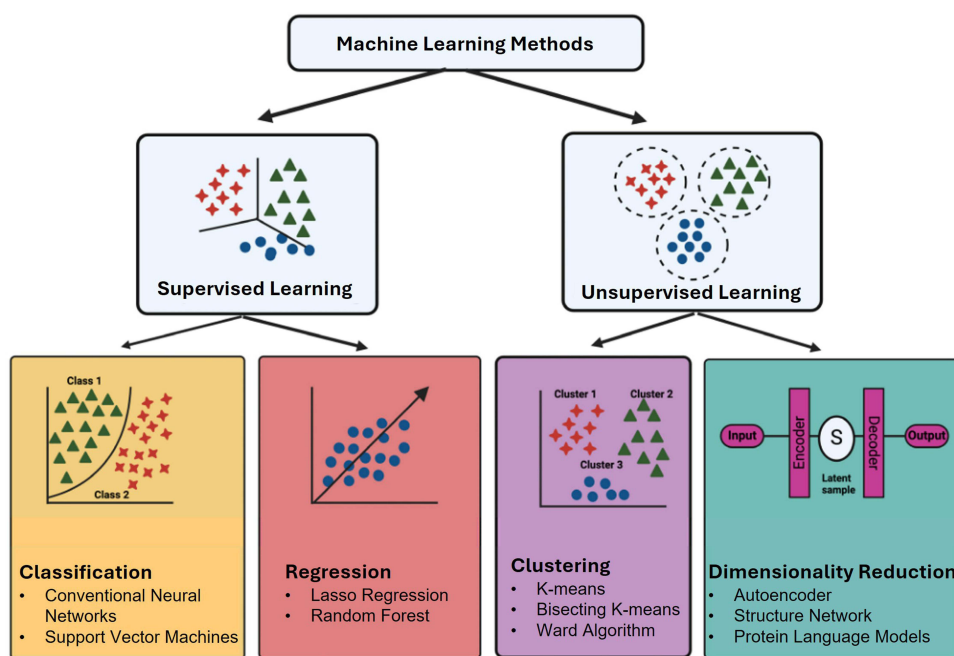


Figure 2 Some of the most frequently implemented techniques and algorithms of supervised and unsupervised learning in drug discovery.⁷⁰

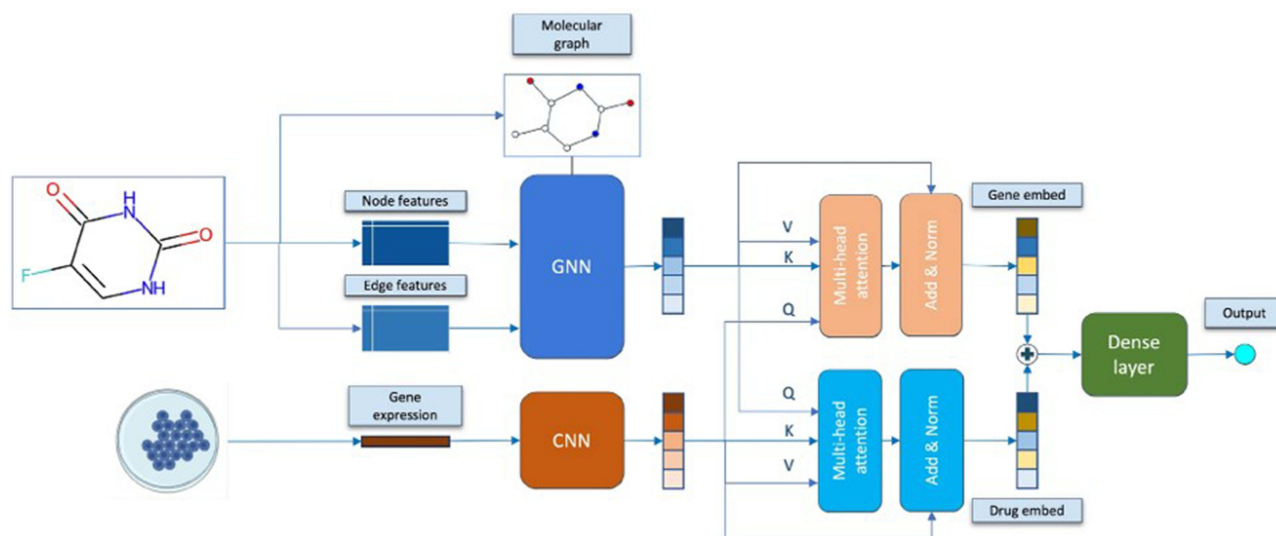


Figure 3 Example deep neural architecture (XGDP) for drug response prediction. A graph neural network encodes molecular structure, a convolutional network encodes cell-line transcriptomics, and an attention module integrates these features.⁷⁵

attention-based saliency highlight that atoms or genes drive predictions.⁷⁵ Overall, deep neural networks, ranging from CNNs for imaging to GNNs for molecules, enable the end-to-end learning of compound screening and molecular profiling. These architectures capture intricate structure-activity relationships (eg, chemical substructures and signaling interactions) that conventional features might miss.⁷⁵ Figure 3 illustrates the GNN and CNN models (XGDP), which exemplify this trend.

Generative Models and Reinforcement Learning

Generative deep learning and RL have emerged as powerful tools for identifying novel candidate molecules and optimizing drug properties.^{79,80} Generative models (VAEs, GANs, normalizing flows, and graph-based generators) can be used to learn the distribution of known anti-cancer compounds and the de novo design of new analogs. For instance,

generative adversarial networks have been trained on breast cancer drugs to produce novel SMILES strings with a predicted potency.⁸¹ Flow-based models generate chemically valid molecules by inverting probabilistic transformations. In these frameworks, the learned latent spaces enable smooth sampling and property optimization. Reinforcement learning, usually combined with generative decoders, guides the generation process toward objectives such as target affinity or toxicity avoidance.⁸² Popular RL-driven frameworks such as REINVENT and Rationale RL use policy gradients to iteratively improve molecules. This model proposes modifications to a seed compound and receives a response from a predictor (eg, docking score).⁸³ Thus, agent-based RL can be used to navigate the chemical reaction space and suggest synthesis routes. For example, a recent application by Zhavoronkov et al used a generative RL algorithm to design DDR1 kinase inhibitors; four candidate molecules were synthesized and two showed potent activity, demonstrating the successful translation of AI-designed leads.⁸³ Moreover, strengths, weaknesses, generalizability, and clinical applicability of various AI and ML based approaches in breast cancer are given in Table 1. Overall, generative models allow hypothesis-free exploration of the chemical space tailored to breast cancer targets (eg, ER and HER2 pathways). They can incorporate condition-based constraints. Conditional VAEs can generate molecules predicted to bind a specific target or modulate a given gene-expression signature. Reward functions in RL pipelines often incorporate predicted bioactivity and drug-likeness and may explicitly optimize ADMET properties to improve clinical viability. These systems have led to new advances in oncology. For example, GAN-based models trained on cancer drug libraries have generated candidate compounds that were retrospectively confirmed as novel ligands.⁸³ The workflow of the generative AI framework for drug design is shown in Figure 4.

Integrative Multi-Omics and Systems Biology

Breast cancer exhibits heterogeneity across the genomic, transcriptomic, proteomic, and metabolomic levels.¹¹ Integrative ML approaches combine these multi-omics layers to improve drug discovery. For example, Malik et al applied supervised feature selection (Neighborhood Component Analysis) to The Cancer Genome Atlas abbreviated as TCGA transcriptomic and proteomic profiles and then used a neural network to predict patient survival and drug-response groups.⁶⁹ This model incorporates regression and k-means clustering to segregate likely responders, outperforming single-omics models. Similarly, Sammut et al combined clinicopathology, DNA mutation, and RNA expression features in ensemble ML models that predicted therapeutic responses with an AUC of 0.87.⁶⁸ Another strategy is late fusion, in which separate models are built for each omics type and combined into a meta-learner. For example, GraphDRP and DeepCDR built graph-based encodings for genomics and proteomics, which were jointly modeled with drug graphs.⁷⁵ These methods extract complementary signals (eg, mutation-driven signaling vs expression subtypes) to highlight patient-specific drug sensitivities. In metabolomics, unsupervised clustering of combined gene/protein/metabolite networks has revealed metabolic vulnerabilities unique to certain breast cancer subpopulations, suggesting the need for targeted drugs. Machine learning also enables pan-omics signatures for repurposing by mapping the transcriptomic perturbation of a drug (LINCS-L1000 data) onto tumor profiles. Algorithms such as Connectivity Map compute similarity scores to suggest existing drugs that reverse cancer signatures.¹⁰⁶ These integrative pipelines often use graphical models or tensor factorizations to reconcile the heterogeneous data types. The ML-driven strategies for breast cancer drug discovery, showing breast cancer-relevant molecular targets, applied ML methodologies, and identified candidate compounds are given in Table 2.

Data Resources and Benchmarks

Modern ML for breast cancer relies on extensive datasets and benchmarks. Public repositories provide genomic, chemical, transcriptomic, and clinical data essential for model training and validation. Large-scale cancer genomics projects (eg, TCGA and METABRIC) provide tumor mutational and expression profiles for thousands of patients.¹¹⁸ Drug discovery leverages chemical libraries (eg, ChEMBL and ZINC) containing millions of compounds with annotated bioactivities. Gene expression sources (GEO and LINCS L1000) archive perturbation signatures for repurposing efforts. Clinical trial registries (ClinicalTrials.gov) catalog patient outcomes and therapeutic regimens. These resources reinforce ML tasks, as genomic datasets enable mutation drug association models,⁶⁸ chemical databases support virtual screening, transcriptomic compendia facilitate drug disease signature matching, and trial data guide response predictions.^{68,118}

Table 1 AI and ML-Based Approaches for Breast Cancer Drug Discovery Research

Method/Approach	Strengths	Weaknesses/Data Biases	Generalizability & Clinical Applicability	References
Regression and DNN QSAR for drug pairs	Strong performance	QSAR may overfit if the descriptors are biased	Useful for hit prioritization; needs external dataset testing	[85]
Integrated network ML (cell-line and drug features)	High predictive correlation across GDSC & CCLE data	Limited in vivo/clinical response training	Promising for in vitro prediction; needs clinical validation	[86]
Deep learning for drug synergy	Captures complex non-linear interactions	Model interpretability challenges	Useful for combination prioritization; requires experimental validation	[87]
ML for ADMET properties	Predicts pharmacokinetic profiles	Dataset biases affect pharmacological metrics	Early stage; aids candidate filtering	[88]
Multi-omics ML framework for response	High AUC; integrates mutation, expression	Requires rich multi-layer clinical data	Potential for personalized prediction; limited public clinical cohorts	[89]
QSAR and regression for TNBC inhibitors	High R ² ; descriptor-level insight	Focused on one subtype (MDA-MB-231)	Good TNBC applicability; needs diverse cell models	[90]
Ensemble and ML for synergistic pairs	Combines feature annotations and network information	Validation limited to cell/3D-TSC models	Preclinical; supports experimental synergy testing	[91]
Overview of DL models for drug response	Comprehensive workflow guidance	General ML constraints (data sparsity)	Framework useful for clinical researchers	[92]
Classification & regression for combo synergy	Systematic synergy modeling	Limited breast cancer-specific figures	Good general combo insights; needs BC-focused sets	[93]
Comparative ML models (RF, SVM, AdaBoost)	Includes SHAP for interpretability	Clinical outcome complexity limits generality	Useful for risk stratification; clinical deployment needs prospective validation	[94]
AI framework for gene expression features	Extracts potential markers linked to drug response	Method requires validation on larger cohorts	Offers biological insight; the translational gap remains	[95]
Review of ML/DL drug response approaches	Broader context: performance summary	Not BC-exclusive but includes key BC studies	Guides model selection; highlights limitations	[96]
MRI and ML for therapy response	Longitudinal data utilization	Complex image preprocessing	High translational value for treatment outcomes	[97]
Interpretable ML for drug response	Focus on interpretability	Performance varies with cohort	May guide clinical decision support	[98]

(Continued)

Table 1 (Continued).

Method/Approach	Strengths	Weaknesses/Data Biases	Generalizability & Clinical Applicability	References
Vector quantization network for synergy	Robust to noisy inputs	Early stage, BC specificity untested	Useful for high-throughput prediction	[99]
ML with imbalance handling	Highlights bias solutions	Not drug discovery-specific	Important general ML limitation	[100]
Imaging-based ML	Diagnostic accuracy improvements	Focused on detection, not drug discovery	Useful data patterns; limited drug context	[101]
XAI in breast cancer	Highlights interpretability methods	General detection focus	Guides trustworthy models, indirectly drug response	[102]
Systematic review of DL risk models	Method comparisons	Focus on risk prediction	Highlights clinical validation gaps	[103]
Perspectives on ML in drug discovery	Connects ML with clinical R&D	Conceptual; not experimental	Informs the AI strategy in breast cancer	[104]
Histopathologic DL review	Model comparisons; interpretability issues	Imaging-first focus	Highlights data generalization issues	[105]

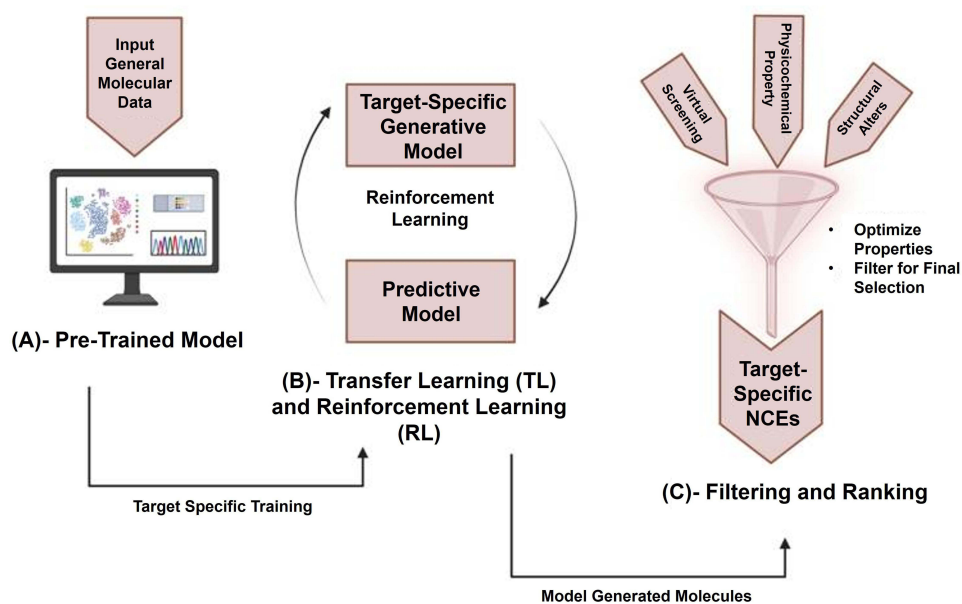


Figure 4 Workflow of a generative AI framework for drug design. (A) Unlabeled chemical databases are used to pretrain a generative model, (B) reinforcement learning then fine-tunes the model with a reward network to generate optimized compounds, and (C) filtering and ranking for shortlisting optimized compounds.⁸⁴

Genomic Data

Genomic repositories provide the backbone for variant-driven drug discovery. The TCGA Breast (BRCA) dataset includes approximately 1,098 tumor samples with comprehensive molecular data (whole-exome mutations, copy number, methylation, and RNA-seq).¹¹⁸ Similarly, METABRIC analyzed approximately 2,000 breast tumors (discovery and

Table 2 ML-Driven Strategies for Breast Cancer Drug Discovery

Targets/Subtype	ML Approach	Candidate Compounds/ Repurposed Drugs	Validation	References
Subtype drug repurposing (including TNBC)	Network-based ML integrating genomic and pathway data	Identified single and paired repurposable drugs for multiple BC subtypes	In silico	[107]
Multiple targets across breast cancer cell lines	Deep neural network predicting drug synergy	Synergistic small-molecule pairs via inhibition features	In silico synergy prediction	[87]
Drug combinations across 52 BC cell lines	Regression ML and Deep learning QSAR	Anchor and library drug pairs (eg, combinational IC ₅₀ modeling)	In silico QSAR with high R ²	[85]
RTK signaling targets	Multimodal deep neural network (MM-DNN) for QSAR	37 structurally similar compounds to existing drugs	In silico QSAR with rigorous validation metrics	[108]
EGFR, HER2, ER, NF-κB, PR	GA-SVM-SVM classification and docking	Identified 4454 ligands with precise target assignments	Computational screening	[109]
Breast cancer signaling pathways	Network medicine and GenAI extraction	108 drugs covering pathways	AI accelerated extraction	[110]
Estrogen receptor modulators (ER+)	Generative AI guided by pharmacophore	Novel candidates with drug-like properties	In silico generative hits	[111]

(Continued)

Table 2 (Continued).

Targets/Subtype	ML Approach	Candidate Compounds/ Repurposed Drugs	Validation	References
ERBB2 (HER2+, target)	RF, SVM, LASSO, ANN, CNN	59 high-affinity compounds predicted	In silico target-compound identification	[112]
Pyroptosis pathway in TNBC	Biofactor-regulated neural network	Pyroptosis-inducing drug pairs (mitoxantrone and gambogic acid)	In vivo TNBC model validated anti-tumor immune effects	[113]
Nanomaterial carrier; 5HZN protein	Reinforcement learning (MoLDQN)	6000 predicted drug molecules	Biological activity in vitro	[114]
Oncology targets across cancers including BC	Protein/network feature ML scoring	Prioritized druggable targets	Validation on clinical trial drug targets	[115]
TNBC subgroups	Patient stratification and repurposing	Top chemotherapy and targeted mechanisms	In silico subgroup stratification	[116]
QSAR repurposing across BC	Deep CNN and transfer learning	Repurposed compounds via QSAR	Computational	[117]

validation cohorts) by integrating somatic copy number and gene expression.^{118,119} These rich datasets have enabled ML models to correlate somatic alterations with drug sensitivity. For instance, integrated ML models using TCGA/METABRIC features can predict chemotherapy response. Sammut et al found that multi-omic input (mutation, copy number, and expression) yielded an AUC of 0.87 in classifying pathologic complete responders.⁶⁸ Additionally, population-scale consortia such as ICGC contribute to further cohorts with genomic data from diverse ancestries.¹¹⁸ ML-driven analyses of these data have identified mutation signatures (eg, BRCA1/2 and PI3K) that inform targeted therapy selection and have been used to build predictive classifiers for drug efficacy. Large germline resources and cancer variant compendia complement these by providing frequency and functional annotations. Altogether, public genomic datasets (eg, TCGA and METABRIC) are crucial for building mutation-drug correlation models that guide targeted treatment choices.^{68,118}

Chemical Data

Chemical databases provide structures and activities for drug modeling. ChEMBL is a manually curated repository of bioactive molecules, linking chemical structures to binding, functional assay, and ADMET data.^{120,121} For example, it aggregates assay results, such as 65,534 anti-cancer assays, across compounds, enabling ML training on ligand-based activities.¹¹⁸ ZINC provides a freely searchable catalog of purchasable small molecules (over 750 million by 2023), often used for virtual screening libraries.^{122,123} These resources allow ML pipelines to learn structure-activity relationships: ligand-based QSAR models use ChEMBL to train predictors of breast cancer cell line viability, and virtual screening workflows dock ZINC compounds against targets. Recent studies have used ChEMBL activity matrices as features for machine-learning synergy and sensitivity prediction. The figure below depicts how the chemical features from these sources were fed into the ML models.

Transcriptomic Data

Transcriptomic repositories are invaluable for signature-based repurposing and mechanistic studies. Gene Expression Omnibus (GEO) archives thousands of breast cancer microarray and RNA-seq studies.¹²⁴ The LINCS L1000 program provides genome-wide mRNA perturbation data for cells treated with drugs (profiled with approximately 20,000 - perturbagens).¹²⁵ These datasets support algorithms such as the Connectivity Map that match drug-induced expression

changes to tumor gene signatures. For instance, ML methods can combine cell-line expression (from LINCS) and tumor expression to predict drug responses or repurposing hits. A recent BMC Genomics study fused tumor transcriptomes with proteomes (from TCGA and other sources) for survival and drug-response prediction, employing feature selection (NCA) and deep neural nets.⁶⁹ Moreover, compendia such as TCGA include matching gene-expression and methylation profiles, enabling the ML fusion of multiple transcriptomic layers. In practice, publicly available signatures (GEO datasets) are used to train and validate treatment response classifiers and to prioritize candidate repurposed drugs by signature reversal.

Clinical Data

Clinical datasets connect molecular predictions to patient outcomes. ClinicalTrials.gov is the largest registry of active and completed oncology trials, including many breast cancer studies.¹²⁶ Data from trials, often genomically annotated, were mined to correlate the biomarkers with therapeutic success. For example, consortium databases and published trial cohorts allow ML models to be trained using real-world response data (eg, by linking PIK3CA mutation status to PI3K inhibitor efficacy). Another source is registry data from hospital networks (eg, AACR Project GENIE) that aggregate genomic profiles with clinical annotations.¹²⁷ These clinical repositories enable supervised learning of predictive models for targeted therapy responses. One study showed that integrating genomic features from trials into ML classifiers could predict responders with high accuracy.⁶⁸ In summary, clinical datasets provide the ground truth for ML validation, enabling the translation of molecular insights into predicted patient benefits.^{68,118} The application of ML to convert various clinical data sources into valuable insights for cancer therapy is illustrated in Figure 5. Various data types, such as electronic health records (EHR), digital pathology slides, radiology images, data from wearable devices, and multi-omics, can be utilized for ML-driven clinical oncology.¹²⁸ Unsupervised clustering, identification of multi-omics subgroups, and deep phenotyping are some of the creative discoveries made possible by AI systems, which also improve productivity by automating tasks, such as classification, differential subtype analysis, treatment response prediction, and regression analyses, such as gene expression scoring. Table 3 summarizes the main datasets and their applications in breast cancer ML research.

Translational Insights and Case Studies

ML has begun to demonstrate real-world effects in breast cancer pharmacotherapy.¹²⁹ One prominent application of AI is in AI-driven drug repurposing. ML models can predict the novel uses of existing drugs by matching their molecular signatures. For example, gene-expression-based ML has identified baicalein (a herbal compound) as a modulator of HIF-1 α that re-sensitizes tamoxifen-resistant breast cancer cells.¹³⁰ This suggests that repurposing agents to overcome endocrine resistance can be guided by an ML-based signature analysis. Likewise, large-scale ML analyses have identified

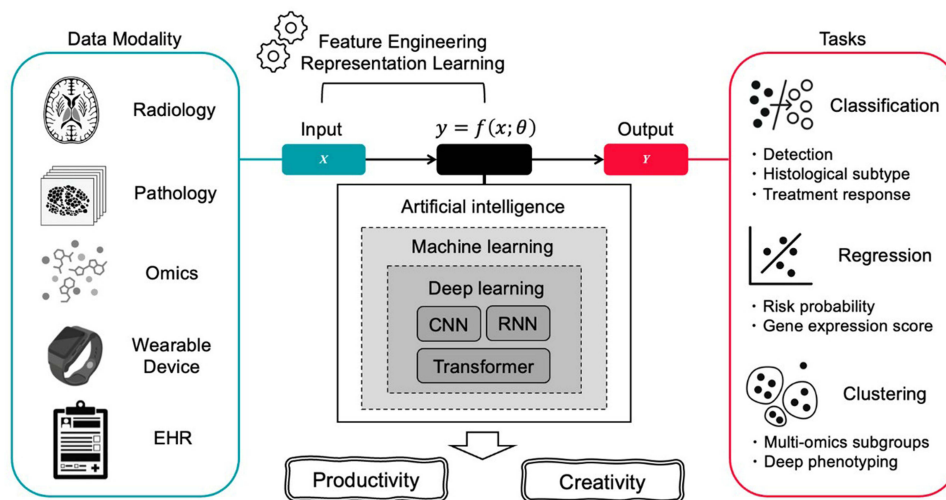


Figure 5 An overview of ML-driven strategies and their relevance to oncology.¹²⁸

Table 3 Main Datasets Supporting ML-Driven Breast Cancer Drug Discovery

Dataset	Primary Data Type	Breast Cancer Relevance	Characteristic ML Use Cases	Translational Value
TCGA-BRCA	Multi-omics (WES, RNA-seq, CNV, methylation), clinical outcomes	Large cohort of primary breast tumors; subtype-resolved (HR+, HER2+, TNBC)	Target discovery, subtype classification, biomarker identification, drug response prediction	Links molecular signatures to patient outcomes; clinical relevance
CCL/DepMap	Genomics, transcriptomics, CRISPR/RNAi dependency, drug sensitivity	Breast cancer cell lines with dependency scores	Target essentiality prediction, synthetic lethality, mechanism inference	Guides target prioritization and vulnerability mapping
METABRIC	Gene expression, copy number, long-term clinical follow-up	Independent validation cohort; robust subtype and survival data	Prognostic modeling, risk stratification, therapy response modeling	Strong outcome annotation; complements TCGA
ChEMBL	Bioactivity, targets, ADMET annotations	Contains ER, HER2, PI3K, CDK4/6 ligand data	QSAR modeling, target-ligand prediction, ML-based screening	Supports lead identification and optimization
GDSC/CTRP	Drug response (IC ₅₀ , AUC) across cancer cell lines	Breast cancer cell line pharmacogenomics	Drug sensitivity prediction, combination modeling, biomarker-drug associations	Preclinical drug response benchmarking
LINCS (LI000)	Gene expression perturbation signatures	Drug- and gene-perturbed breast cancer cell lines	Drug repurposing, mechanism of action inference, signature matching	Enables connectivity-map-based repurposing
AACR GENIE	Clinical-grade genomic sequencing + treatment data	Advanced/metastatic breast cancer cases	Real-world evidence modeling, mutation-drug associations	Bridges ML predictions with clinical practice
ZINC	Virtual chemical libraries (drug-like compounds)	Source of candidate molecules for BC targets	Virtual screening, generative ML, docking-ML pipelines	Expands chemical search space for novel candidates
GEO	Transcriptomics, epigenomics, functional genomics	Large number of breast cancer expression studies	Signature discovery, drug response stratification, model training	Cost-effective access to diverse patient cohorts

repurposed kinase inhibitors for HER2-positive or triple-negative subtypes by linking drug-target networks with tumor omics. Another area of research is biomarker-guided therapies. ML classifiers trained in genomic features have helped to refine patient stratification. Sammut et al used an ensemble ML model combining DNA mutations and RNA expression to predict which HER2+ patients would attain a complete response to neoadjuvant therapy, thus enabling more precise treatment planning.⁶⁸ In practice, such models can complement established markers (eg, ER, HER2, and PIK3CA) to tailor their combinations. For example, interpretable RF models have been developed to predict the synergistic effects of MEK/PI3K inhibitor pairs, revealing that NF- κ B and mTORC1 modulate resistance in specific breast cancer cell-lines. These models suggest a combination of regimens for resistant tumors and illustrate how ML can uncover the hidden mechanisms. Industrial pipelines showcase the translational potential of AI. In silico Medicine (founded by Zhavoronkov) has successfully incorporated AI-designed compounds into the clinic. Their reinforced GAN platform generated DDR1 inhibitors, which were further validated *in vitro*.^{83,131} AstraZeneca has publicly invested in ML for immuno-oncology and small-molecule design (although the detailed pipelines are proprietary). BenevolentAI collaborates with pharma to mine biomedical literature and genomics; in breast cancer, their knowledge graph approaches prioritized repurposing candidates for subtypes.

Another example is the deep-learning platform DeepMatter/Exscientia, which discovered potent A2A receptor antagonists (anti-cancer immunotherapy) in Phase I trials. Overall, these case studies illustrate the potential of ML to

shorten the hit-to-lead cycle. Generative models quickly propose drug-like molecules, predictive models validate them in silico, and some AI-suggested compounds have progressed to experimental testing.^{83,130} Importantly, retrospectives on ML models have highlighted the interpretability gains. For instance, the RF synergy model mentioned above provides direct insights into protein-drug interactions in HER2+ cells, which can inform wet-lab hypothesis generation. Many of the cited industry successes (DDR1 by in silico and A2A antagonist by Exscientia) were accompanied by peer-reviewed publications that detailed the ML pipeline and biological validation, underscoring that AI approaches can yield actionable therapeutic candidates.

Challenges and Limitations

Despite the progress, major obstacles have slowed the clinical translation of ML-based discoveries. Data-related issues are paramount, in which the available datasets are often small, noisy, and imbalanced.¹³² For breast cancer, drug response data (particularly for combination therapies) are limited. This makes the models prone to overfitting.^{24,133} Even large groups such as TCGA have missing data modalities and batch effects. Noisy experimental measurements (eg, variable IC₅₀ assays) lead to hallucinations or unreliable predictions.^{24,134} Consequently, models risk poor generalizability, as one study warns, neglecting rigorous cross validation with scaffold splits produces overly optimistic estimates.^{24,133} The model-related challenges include interpretability and reproducibility. Deep black-box networks can achieve accuracy, but often reduce transparency, which makes it difficult to trust predictions in a clinical setting.²⁴ Regulatory frameworks demand a mechanistic understanding, or at least confidence, in model reasoning. Thus, explainable AI (XAI) techniques are required to justify ML-driven drug hypotheses. Furthermore, the chemical novelty of numerous ML designs raises concerns regarding their patentability and synthetic feasibility.

At the system level, clinical translation enforces non-technical restrictions. Regulatory and ethical issues, such as sharing patient molecular data, are hampered by privacy laws.²⁴ Bias in the training data (eg, under-representation of minority groups) can lead to inequitable predictions. Moreover, there is always a gap between computational predictions and biological validation. Many ML studies end with in silico results without experimental follow-ups. Bridging this gap requires expensive assays and interdisciplinary collaboration. Only approximately 10–15% of predicted drug leads succeed in clinical trials,^{69,135,136} which reflect the complexity of human biology and pharmacology. One study reported that this low success rate suggests limitations in our understanding of cancer complexity and emphasizes the need for better modeling strategies.^{69,136} Despite these challenges, the field has been actively evolving. Federated learning and data amplification may increase data scarcity. Explainable models and integrated omics aim to increase trust. Policy efforts are beginning to create guidelines for AI in drug development. However, significant investment is required in data curation, rigorous validation, and cross-disciplinary infrastructure to realize the potential of ML in clinics.

Future Directions

In the future, emerging ML patterns promise to further transform drug discovery in breast cancer.¹³² Collective knowledge allows models to train on distributed patient data without data sharing, preserving privacy while increasing the sample size.²⁴ For example, without moving raw genomes, collaborative networks can produce pan-cancer predictors by combining local models on whole exome sequencing. Explainable AI (XAI) methods, such as attention visualization or causal inference, are tailored for biology so that clinicians can understand ML-generated hypotheses.^{137,138} XAI will help to satisfy regulatory requirements and build trust in AI recommendations.¹³⁷ Quantum machine learning has been to accelerate drug screening.^{139,140} Quantum-enhanced algorithms (quantum annealing or variational circuits) can navigate chemical space faster or model molecular simulations more accurately.²⁴ Beyond classical capabilities, quantum approaches have been initiated to solve small-molecule optimization problems.

Advances in reinforcement learning can spawn multi agent systems for integrated drug design and synthesis planning.^{141,142} In such a situation, one agent might propose novel scaffold modifications, whereas another agent schedules synthetic steps. Both of these works cooperatively to optimize efficiency and manufacturing. Similarly, digital twins and systems pharmacology models of individual patients are likely to play a role. In silico patient avatars that simulate tumor growth and drug response can personalize regimen selection.¹⁴³ This concept is already being used in oncology, where ML-driven simulations of drug pharmacokinetics and tumor evolution support adaptive therapy design.

Digital twins are a virtual representation of a tumor of individual patient that evolves over time as new data are assimilated. In oncology, digital twin frameworks use machine learning and systems biology to simulate tumor growth, treatment responses, and personalized therapeutic regimens. Digital twin concepts are especially compelling for triple-negative breast cancer (TNBC), where treatment optimization remains challenging. A recent study constructed MRI-based digital twins for TNBC patients to model and optimize neoadjuvant chemotherapy schedules, achieving better predicted pathological complete response rates through *in silico* exploration of treatment schedules.¹⁴⁴ Mechanistically, oncology digital twins integrate clinical imaging, molecular profile, and treatment data into predictive models that can be iteratively updated as patient data accrue. Reviews highlight that such twins can simulate disease progression, forecast treatment response, and explore hypothetical interventions before clinical application, all while incorporating patient-specific molecular and physiological features.^{145,146} For drug discovery, virtual clinical trials and *in silico* simulations anchored in breast cancer digital twins could compare candidate molecules, optimize dosing regimens, and simulate response heterogeneity across subtypes. Preliminary work in other cancers utilizing TCGA, CCLE, and CTRP data shows that digital twin simulations can recapitulate historical clinical outcomes, implying an analogous application in breast cancer drug response modeling.

Conclusion

ML has become an increasingly influential component of breast cancer discovery, enabling more efficient target identification, compound prioritization, and mechanistic insight through the integration of large-scale molecular and clinical datasets. As discussed in this review, advances in supervised and unsupervised learning, deep neural networks, and generative models have expanded the capacity to model complex structure activity relationships, uncover subtype-specific vulnerabilities, and support data-driven drug repurposing strategies. These developments reflect substantial methodological progress and the growing availability of high-quality biomedical resources. However, significant challenges continue to limit the translational impact of AI-driven approaches. Many studies rely on retrospective or highly curated datasets that inadequately represent breast cancer heterogeneity, leading to concerns regarding bias, reproducibility, and generalizability. In addition, the widespread use of opaque deep learning models complicates biological interpretation, regulatory approval, and clinical adoption. Crucially, relatively few AI-derived drug candidates have advanced beyond preclinical validation, underscoring the need for tighter coupling between computational prediction, experimental, and clinical evidence. The future research should focus interpretable, robust, and uncertainty-aware models that are evaluated across independent cohorts and breast cancer subtypes. Greater emphasis on integrative frameworks combining multi-omics data, functional validation platforms, and real-world clinical data will be essential for meaningful clinical translation. Emerging directions such as digital twin modeling, federated learning, and ML-guided trial design hold promise when grounded in breast cancer-specific data and realistic validation strategies. Overall, while AI-based drug discovery is not a panacea, its thoughtful integration into biologically and clinically informed workflows has the potential to substantially improve the development of personalized therapies for breast cancer.

Author Contributions

All authors made a significant contribution to the work reported, whether that is in the conception, study design, execution, acquisition of data, analysis and interpretation, or in all these areas; took part in drafting, revising or critically reviewing the article; gave final approval of the version to be published; have agreed on the journal to which the article has been submitted; and agree to be accountable for all aspects of the work.

Funding

This study did not receive any external funding.

Disclosure

The authors declare no conflict of interest.

References

- Zhou J, Guo Z, Peng X, et al. Chrysotoxine regulates ferroptosis and the PI3K/AKT/mTOR pathway to prevent cervical cancer. *J Ethnopharmacol.* 2025;338:119126. doi:10.1016/j.jep.2024.119126
- Wang ZB, Zhang X, Fang C, et al. Immunotherapy and the ovarian cancer microenvironment: exploring potential strategies for enhanced treatment efficacy. *Immunology.* 2024;173(1):14–32. doi:10.1111/imm.13793
- Peng X, Liu Z, Luo C, et al. PSMD12 promotes hepatocellular carcinoma progression by stabilizing CDK1. *Front Immunol.* 2025;16:1581398. doi:10.3389/fimmu.2025.1581398
- Bernal G, Aquea G, Ramirez-Rivera S. Metal-based molecules in the treatment of cancer: from bench to bedside. *Oncology Research.* 2025;33(4):759. doi:10.32604/or.2024.057019
- Ma X, Cheng H, Hou J, et al. Detection of breast cancer based on novel porous silicon Bragg reflector surface-enhanced Raman spectroscopy-active structure. *Chin Opt Lett.* 2020;18(5):051701. doi:10.3788/COL202018.051701
- Kim J, Harper A, McCormack V, et al. Global patterns and trends in breast cancer incidence and mortality across 185 countries. *Nat Med.* 2025;2205:1–9.
- Zhou L, Ou S, Liang T, et al. MAEL facilitates metabolic reprogramming and breast cancer progression by promoting the degradation of citrate synthase and fumarate hydratase via chaperone-mediated autophagy. *FEBS J.* 2023;290(14):3614–3628. doi:10.1111/febs.16768
- WHO. Breast Cancer. 2025. Available from: <https://www.who.int/news-room/fact-sheets/detail/breast-cancer#:~:text=,out%20of%20185%20in%202022>. August 14, 2025.
- Han D, Wang T, Li X, et al. Small extracellular vesicles orchestrated pathological communications between breast cancer cells and cardiomyocytes as a novel mechanism exacerbating anthracycline cardiotoxicity by fueling ferroptosis. *Redox Biol.* 2025;86:103843. doi:10.1016/j.redox.2025.103843
- Sha R, Kong X-M, Li X-Y, Wang Y-B. Global burden of breast cancer and attributable risk factors in 204 countries and territories, from 1990 to 2021: results from the Global Burden of Disease Study 2021. *Biomarker Res.* 2024;12(1):87. doi:10.1186/s40364-024-00631-8
- Fumagalli C, Barberis M. Breast cancer heterogeneity. *Diagnostics.* 2021;11(9):1555. doi:10.3390/diagnostics11091555
- Orrantia-Borunda E, Anchondo-Nuñez P, Acuña-Aguilar LE, Gómez-Valles FO, Ramírez-Valdespino CA. Subtypes of breast cancer. *Breast Cancer.* 2022;2022:1.
- Li X, Xiang J, Wang J, Li J, Wu F-X, Li M. FUNMarker: fusion network-based method to identify prognostic and heterogeneous breast cancer biomarkers. *IEEE/ACM Transactions on Computational Biology and Bioinformatics.* 2020;18(6):2483–2491. doi:10.1109/TCBB.2020.2973148
- De Vita A, Liverani C, Molinaro R, et al. Lysyl oxidase engineered lipid nanovesicles for the treatment of triple negative breast cancer. *Sci Rep.* 2021;11(1):5107. doi:10.1038/s41598-021-84492-3
- Molinaro R, Martinez JO, Zinger A, et al. Leukocyte-mimicking nanovesicles for effective doxorubicin delivery to treat breast cancer and melanoma. *Biomater Sci.* 2020;8(1):333–341. doi:10.1039/C9BM01766F
- Nicolini A, Ferrari P. Targeted therapies and drug resistance in advanced breast cancer, alternative strategies and the way beyond. *Cancers.* 2024;16(2):466. doi:10.3390/cancers16020466
- Bobo D, Robinson KJ, Islam J, Thurecht KJ, Corrie SR. *Nanoparticle-Based Medicines: A Review of FDA-Approved Materials and Clinical Trials to Date.* 2016.
- Fornaguera C, García-Celma MJ. Personalized nanomedicine: a revolution at the nanoscale. *J Personalized Med.* 2017;7(4):12. doi:10.3390/jpm7040012
- Jentsch V, Osipenko L, Scannell JW, Hickman JA. Costs and causes of oncology drug attrition with the example of insulin-like growth factor-1 receptor inhibitors. *JAMA Network Open.* 2023;6(7):e2324977. doi:10.1001/jamanetworkopen.2023.24977
- Stock JK, Jones NP, Hammonds T, Roffey J, Dillon C. Addressing the right targets in oncology: challenges and alternative approaches. *J Biomol Screen.* 2015;20(3):305–317. doi:10.1177/1087057114564349
- Di Leo A, Curigliano G, Diéras V, et al. New approaches for improving outcomes in breast cancer in Europe. *Breast.* 2015;24(4):321–330. doi:10.1016/j.breast.2015.03.001
- Baselga J, Im S-A, Iwata H, et al. Buparlisib plus fulvestrant versus placebo plus fulvestrant in postmenopausal, hormone receptor-positive, HER2-negative, advanced breast cancer (BELLE-2): a randomised, double-blind, placebo-controlled, Phase 3 trial. *Lancet Oncol.* 2017;18(7):904–916. doi:10.1016/S1470-2045(17)30376-5
- Pandey V, Verma P, Natesan G, Chaube U. Advancements in breast cancer therapy: integrating AI tools for drug discovery and clinical trials. *Comput Biol Med.* 2025;197:111067. doi:10.1016/j.combiomed.2025.111067
- Sutanto H, Fetarayani D. Integrating artificial intelligence into small molecule development for precision cancer immunomodulation therapy. *Npj Drug Discovery.* 2025;2(1):25. doi:10.1038/s44386-025-00029-y
- Botlagunta M, Botlagunta M, Venkata MD, Kanakapudi C, Khan Z. Correlation-Based Comparative Machine Learning Analysis for the Classification of Metastatic Breast Cancer Using Blood Profile. *Eurasian J Med Oncol.* 2024;8(2):152–164.
- Wang J, Zeng Z, Li Z, et al. The clinical application of artificial intelligence in cancer precision treatment. *J Transl Med.* 2025;23(1):120. doi:10.1186/s12967-025-06139-5
- Hondelink LM, Hüyük M, Postmus PE, et al. Development and validation of a supervised deep learning algorithm for automated whole-slide programmed death-ligand 1 tumour proportion score assessment in non-small cell lung cancer. *Histopathology.* 2022;80(4):635–647. doi:10.1111/his.14571
- Shamai G, Livne A, Polónia A, et al. Deep learning-based image analysis predicts PD-L1 status from H&E-stained histopathology images in breast cancer. *Nat Commun.* 2022;13(1):6753. doi:10.1038/s41467-022-34275-9
- Mu W, Jiang L, Shi Y, et al. Non-invasive measurement of PD-L1 status and prediction of immunotherapy response using deep learning of PET/CT images. *Journal for Immunotherapy of Cancer.* 2021;9(6):e002118. doi:10.1136/jitc-2020-002118
- DiMasi JA, Grabowski HG, Hansen RW. Innovation in the pharmaceutical industry: new estimates of R&D costs. *Journal of Health Economics.* 2016;47:20–33. doi:10.1016/j.jhealeco.2016.01.012
- Haines DD, Rose SC, Cowan FM, Mahmoud FF, Rizvanov AA, Tosaki A. Leveraging Artificial Intelligence and Modulation of Oxidative Stressors to Enhance Healthspan and Radical Longevity. *Biomolecules.* 2025;15(11):1501. doi:10.3390/biom15111501

32. Vamathevan J, Clark D, Czodrowski P, et al. Applications of machine learning in drug discovery and development. *Nat Rev Drug Discov.* 2019;18(6):463–477. doi:10.1038/s41573-019-0024-5
33. Rakhshaninejad M, Fathian M, Shirkoohi R, Barzinpour F, Gandomi AH. Refining breast cancer biomarker discovery and drug targeting through an advanced data-driven approach. *BMC Bioinf.* 2024;25(1):33. doi:10.1186/s12859-024-05657-1
34. Wu Y, Han Y, Li Q, et al. Predictive value of topoisomerase II alpha protein for clinicopathological characteristics and prognosis in early breast cancer. *Breast Cancer Res Treat.* 2022;193(2):381–392. doi:10.1007/s10549-022-06559-7
35. You Y, Lai X, Pan Y, et al. Artificial intelligence in cancer target identification and drug discovery. *Signal Transduct Target Ther.* 2022;7(1):156. doi:10.1038/s41392-022-00994-0
36. Taylor IW, Linding R, Warde-Farley D, et al. Dynamic modularity in protein interaction networks predicts breast cancer outcome. *Nature Biotechnol.* 2009;27(2):199–204. doi:10.1038/nbt.1522
37. Raman K, Damaraju N, Joshi GK. The organisational structure of protein networks: revisiting the centrality–lethality hypothesis. *Systems Synthet Biol.* 2014;8(1):73–81. doi:10.1007/s11693-013-9123-5
38. Hao D, Li C. The dichotomy in degree correlation of biological networks. *PLoS One.* 2011;6(12):e28322. doi:10.1371/journal.pone.0028322
39. Carson MB, Lu H. Network-based prediction and knowledge mining of disease genes. *BMC Med. Genomics.* 2015;8(Suppl 2):S9. doi:10.1186/1755-8794-8-S2-S9
40. Chen K-H, Wang K-J, Wang K-M, Angelia M-A. Applying particle swarm optimization-based decision tree classifier for cancer classification on gene expression data. *Appl Soft Comput.* 2014;24:773–780. doi:10.1016/j.asoc.2014.08.032
41. Zhang L, Liu G, Kong M, et al. Revealing dynamic regulations and the related key proteins of myeloma-initiating cells by integrating experimental data into a systems biological model. *Bioinformatics.* 2021;37(11):1554–1561. doi:10.1093/bioinformatics/btz542
42. LeCun Y, Bengio Y, Hinton G. Deep learning. *Nature.* 2015;521(7553):436–444. doi:10.1038/nature14539
43. Hao X, Zhang G, Ma S. Deep learning. *Int J Semantic Comput.* 2016;10(03):417–439. doi:10.1142/S1793351X16500045
44. Cornu H. Machine learning to identify and prioritise drug targets. 2023. Available from: <https://www.embl.org/news/science/machine-learning-to-identify-and-prioritise-drug-targets/#:~:text=To%20address%20this%20challenge%2C%20Open,main%20source%20of%20common%20genetic>. Accessed Feb 13, 2026.
45. Wu X, Zahari MS, Renuse S, et al. Phosphoproteomic analysis identifies focal adhesion kinase 2 (FAK2) as a potential therapeutic target for tamoxifen resistance in breast cancer. *Mol Cell Proteomics.* 2015;14(11):2887–2900. doi:10.1074/mcp.M115.050484
46. Gentile F, Yaacoub JC, Gleave J, et al. Artificial intelligence-enabled virtual screening of ultra-large chemical libraries with deep docking. *Nature Protocols.* 2022;17(3):672–697. doi:10.1038/s41596-021-00659-2
47. Caba K, Tran-Nguyen V-K, Rahman T, Ballester PJ. Comprehensive machine learning boosts structure-based virtual screening for PARP1 inhibitors. *J Cheminf.* 2024;16(1):40. doi:10.1186/s13321-024-00832-1
48. Shafiq M, Sherwani ZA, Mushtaq M, Nur-e-Alam M, Ahmad A, Ul-Haq Z. A deep learning-based theoretical protocol to identify potentially isoform-selective PI3Kα inhibitors. *Mol Diversity.* 2024;28(4):1907–1924. doi:10.1007/s11030-023-10799-0
49. Jaiswal R, Bhati G, Ahmed S, Siddiqi MI. iDCNNPred: an interpretable deep learning model for virtual screening and identification of PI3K α inhibitors against triple-negative breast cancer. *Mol Diversity.* 2025;29(4):3077–3100. doi:10.1007/s11030-024-11055-9
50. Mahanta HJ, Boruah A, Phukan B, Chutia H, Bharali P, Nagamani S. Exploring graph-based models for predicting active compounds against triple-negative breast cancer. *Mol Diversity.* 2025;29(1):1–19. doi:10.1007/s11030-024-10829-5
51. Noor F, Junaid M, Almalki AH, Almaghrabi M, Ghazanfar S, Tahir UI Qamar M. Deep learning pipeline for accelerating virtual screening in drug discovery. *Sci Rep.* 2024;14(1):28321. doi:10.1038/s41598-024-79799-w
52. De Carlo A, Ronchi D, Piastra M, Tosca EM, Magni P. Predicting ADMET properties from molecule smile: a bottom-up approach using attention-based graph neural networks. *Pharmaceutics.* 2024;16(6):776. doi:10.3390/pharmaceutics16060776
53. Cook D, Brown D, Alexander R, et al. Lessons learned from the fate of AstraZeneca’s drug pipeline: a five-dimensional framework. *Nat Rev Drug Discov.* 2014;13(6):419–431. doi:10.1038/nrd4309
54. Tosca EM, Bartolucci R, Magni P. Application of artificial neural networks to predict the intrinsic solubility of drug-like molecules. *Pharmaceutics.* 2021;13(7):1101. doi:10.3390/pharmaceutics13071101
55. Wei M, Zhang X, Pan X, et al. HobPre: accurate prediction of human oral bioavailability for small molecules. *J Cheminf.* 2022;14(1):1. doi:10.1186/s13321-021-00580-6
56. Berthelsen R, Sjögren E, Jacobsen J, et al. Combining in vitro and in silico methods for better prediction of surfactant effects on the absorption of poorly water soluble drugs—A fenofibrate case example. *Int J Pharm.* 2014;473(1–2):356–365. doi:10.1016/j.ijpharm.2014.06.060
57. Patel CN, Kumar SP, Rawal RM, Patel DP, Gonzalez FJ, Pandya HA. A multiparametric organ toxicity predictor for drug discovery. *Toxicol Mech Methods.* 2020;30(3):159–166. doi:10.1080/15376516.2019.1681044
58. Vijara N, Toopradab B, Yahuafai J, et al. Machine Learning-Driven QSAR Modeling of Anti-Cancer Activity from a Rationally Designed Synthetic Flavone Library. *ChemMedChem.* 2025;2025(15):e202500143. doi:10.1002/cmdc.202500143
59. Li Y, Xu Y, Yu Y. CRNNTL: convolutional recurrent neural network and transfer learning for QSAR modeling in organic drug and material discovery. *Molecules.* 2021;26(23):7257. doi:10.3390/molecules26237257
60. Jeon W, Kim D. Autonomous molecule generation using reinforcement learning and docking to develop potential novel inhibitors. *Sci Rep.* 2020;10(1):22104. doi:10.1038/s41598-020-78537-2
61. Venkataraman M, Rao GC, Madavareddi JK, Maddi SR. Leveraging machine learning models in evaluating ADMET properties for drug discovery and development. *ADMET DMPK.* 2025;13(3). doi:10.5599/admet.2772
62. Lee H, Kim J, Kim J-W, Lee Y. Recent advances in AI-based toxicity prediction for drug discovery. *Front Chem.* 2025;13:1632046. doi:10.3389/fchem.2025.1632046
63. Jaume G, de Brot S, Song AH, et al. Deep learning-based modeling for preclinical drug safety assessment. *bioRxiv.* 2024.
64. Chen X, Roberts R, Tong W, Liu Z. Tox-GAN: an artificial intelligence approach alternative to animal studies—a case study with toxicogenomics. *Toxicol Sci.* 2022;186(2):242–259. doi:10.1093/toxsci/kfab157
65. Benzekry S, Mastri M, Nicolò C, Ebos JM. Machine-learning and mechanistic modeling of metastatic breast cancer after neoadjuvant treatment. *PLOS Comput Biol.* 2024;20(5):e1012088. doi:10.1371/journal.pcbi.1012088

66. Gallagher K, Strobl MA, Park DS, et al. Mathematical model-driven deep learning enables personalized adaptive therapy. *Cancer Res.* 2024;84(11):1929–1941. doi:10.1158/0008-5472.CAN-23-2040
67. Ran D, Li J, Zhao M, Du L, Zhang Y, Zhu J. Artificial intelligence integrates multi-omics data for precision stratification and drug resistance prediction in breast cancer. *Front Oncol.* 2025;15:1612474. doi:10.3389/fonc.2025.1612474
68. Sammut S-J, Crispin-Ortuzar M, Chin S-F, et al. Multi-omic machine learning predictor of breast cancer therapy response. *Nature.* 2022;601(7894):623–629. doi:10.1038/s41586-021-04278-5
69. Malik V, Kalakoti Y, Sundar D. Deep learning assisted multi-omics integration for survival and drug-response prediction in breast cancer. *BMC Genomics.* 2021;22(1):214. doi:10.1186/s12864-021-07524-2
70. Patne AY, Dhulipala SM, Lawless W, Prakash S, Mohapatra SS, Mohapatra S. Drug discovery in the age of artificial intelligence: transformative target-based approaches. *Int J Mol Sci.* 2024;25(22):12233. doi:10.3390/ijms252212233
71. Chen H, Engkvist O, Wang Y, Olivecrona M, Blaschke T. The rise of deep learning in drug discovery. *Drug Discovery Today.* 2018;23(6):1241–1250. doi:10.1016/j.drudis.2018.01.039
72. Heenaye-Mamode Khan M, Boodoo-Jahangeer N, Dullull W, et al. Multi-class classification of breast cancer abnormalities using Deep Convolutional Neural Network (CNN). *PLoS One.* 2021;16(8):e0256500. doi:10.1371/journal.pone.0256500
73. Chougrad H, Zouaki H, Alheyane O. Deep convolutional neural networks for breast cancer screening. *Comput Methods Programs Biomed.* 2018;157:19–30. doi:10.1016/j.cmpb.2018.01.011
74. Huo X, Xu J, Xu M, Chen H. An improved 3D quantitative structure-activity relationships (QSAR) of molecules with CNN-based partial least squares model. *Artificial Intelligence the Life Sciences.* 2023;3:100065. doi:10.1016/j.aillsci.2023.100065
75. Wang C, Kumar GA, Rajapakse JC. Drug discovery and mechanism prediction with explainable graph neural networks. *Sci Rep.* 2025;15(1):179. doi:10.1038/s41598-024-83090-3
76. Zuo Z, Wang P, Chen X, Tian L, Ge H, Qian D. SWnet: a deep learning model for drug response prediction from cancer genomic signatures and compound chemical structures. *BMC Bioinf.* 2021;22(1):434. doi:10.1186/s12859-021-04352-9
77. Ramsundar B. *Molecular Machine Learning with DeepChem.* Stanford University; 2018.
78. Zhan K, Lu Z, Zhang Y, editors. Performance optimization for feature extraction section of DeepChem. International Conference on Algorithms and Architectures for Parallel Processing; 2020: Springer.
79. Foster D. *Generative Deep Learning: O'Reilly Media, Inc.* 2022.
80. Arulkumaran K, Deisenroth MP, Brundage M, Bharath AA. Deep reinforcement learning: a brief survey. *IEEE Signal Process Mag.* 2017;34(6):26–38. doi:10.1109/MSP.2017.2743240
81. Das D, Chakrabarty B, Srinivasan R, Roy A. Gex2SGen: designing drug-like molecules from desired gene expression signatures. *J Chem Inf Model.* 2023;63(7):1882–1893. doi:10.1021/acs.jcim.2c01301
82. Pereira T, Abbasi M, Ribeiro B, Arrais JP. Diversity oriented deep reinforcement learning for targeted molecule generation. *J Cheminf.* 2021;13(1):21. doi:10.1186/s13321-021-00498-z
83. Korshunova M, Huang N, Capuzzi S, et al. Generative and reinforcement learning approaches for the automated de novo design of bioactive compounds. *Commun Chem.* 2022;5(1):129. doi:10.1038/s42004-022-00733-0
84. Albani FG, Alghamdi SS, Almutairi MM, Alqahtani T. Artificial Intelligence-Driven Innovations in Oncology Drug Discovery: transforming Traditional Pipelines and Enhancing Drug Design. *Drug Des Devel Ther.* 2025;Volume 19:5685–5707. doi:10.2147/DDDT.S509769
85. Karampuri A, Perugu S. A breast cancer-specific combinatorial QSAR model development using machine learning and deep learning approaches. *Front Bioinform.* 2024;3:1328262. doi:10.3389/fbinf.2023.1328262
86. Huang S, Hu P, Lakowski TM. Predicting breast cancer drug response using a multiple-layer cell line drug response network model. *BMC Cancer.* 2021;21(1):648. doi:10.1186/s12885-021-08359-6
87. Srithanyarat T, Taoma K, Sutthibutpong T, Ruengjitchachawalya M, Liangruksa M, Laomettachtit T. Interpreting drug synergy in breast cancer with deep learning using target-protein inhibition profiles. *BioData Mining.* 2024;17(1):8. doi:10.1186/s13040-024-00359-z
88. Zhou Y, Ye Y, Fan C, editors. Machine Learning-based Algorithm for Screening Drug Candidates in Breast Cancer Treatment. Proceedings of the 2024 5th International Symposium on Artificial Intelligence for Medicine Science; 2024.
89. Rashid MM, Selvarajoo K. Advancing drug-response prediction using multi-modal and-omics machine learning integration (MOMLIN): a case study on breast cancer clinical data. *Briefings Bioinf.* 2024;25(4):bbae300. doi:10.1093/bib/bbae300
90. Khan S, Sarfraz A, Prakash O, Khan F. Machine learning-based QSAR modeling, molecular docking, dynamics simulation studies for cytotoxicity prediction in MDA-MB231 triple-negative breast cancer cell line. *J Mol Struct.* 2024;1315:138807. doi:10.1016/j.molstruc.2024.138807
91. Zhou J-B, Tang D, He L, et al. Machine learning model for anti-cancer drug combinations: analysis, prediction, and validation. *Pharmacol Res.* 2023;194:106830. doi:10.1016/j.phrs.2023.106830
92. Partin A, Brettin TS, Zhu Y, et al. Deep learning methods for drug response prediction in cancer: predominant and emerging trends. *Front Med.* 2023;10:1086097. doi:10.3389/fmed.2023.1086097
93. Abd El-Hafeez T, Shams MY, Elshaier YA, Farghaly HM, Hassanien AE. Harnessing machine learning to find synergistic combinations for FDA-approved cancer drugs. *Sci Rep.* 2024;14(1):2428. doi:10.1038/s41598-024-52814-w
94. Zuo D, Yang L, Jin Y, Qi H, Liu Y, Ren L. Machine learning-based models for the prediction of breast cancer recurrence risk. *BMC Med Inf Decis Making.* 2023;23(1):276. doi:10.1186/s12911-023-02377-z
95. Turki T, Taguchi Y. maGENEgerZ: an Efficient Artificial Intelligence-Based Framework Can Extract More Expressed Genes and Biological Insights Underlying Breast Cancer Drug Response Mechanism. *Mathematics.* 2024;12(10):1536. doi:10.3390/math12101536
96. Singh DP, Kaushik B. A systematic literature review for the prediction of anticancer drug response using various machine-learning and deep-learning techniques. *Chem Biol Drug Des.* 2023;101(1):175–194. doi:10.1111/cbdd.14164
97. Ravi R, Li R, Abdelfatah T, Chan S, Chen X. Breast Cancer Neoadjuvant Chemotherapy Treatment Response Prediction Using Aligned Longitudinal MRI and Clinical Data. arXiv preprint arXiv:251217759. 2025.
98. Ogunleye AZ, Piyawajanusorn C, Gonçalves A, Ghislat G, Ballester PJ. Interpretable machine learning models to predict the resistance of breast cancer patients to doxorubicin from their microRNA profiles. *Adv Sci.* 2022;9(24):2201501. doi:10.1002/advs.202201501
99. Wu J, Yan M, Liu D. Vqsynery: robust drug synergy prediction with vector quantization mechanism. arXiv preprint arXiv:240303089. 2024.

100. Ghavidel A, Pazos P. Machine learning (ML) techniques to predict breast cancer in imbalanced datasets: a systematic review. *J Cancer Survivorship*. 2025;19(1):270–294. doi:10.1007/s11764-023-01465-3
101. Safdar S, Rizwan M, Gadekallu TR, et al. Bio-imaging-based machine learning algorithm for breast cancer detection. *Diagnostics*. 2022;12(5):1134. doi:10.3390/diagnostics12051134
102. Ghasemi A, Hashtarkhani S, Schwartz DL, Shaban-Nejad A. Explainable artificial intelligence in breast cancer detection and risk prediction: a systematic scoping review. *Cancer Innovation*. 2024;3(5):e136. doi:10.1002/cai2.136
103. Hussain S, Ali M, Naseem U, et al. Breast cancer risk prediction using machine learning: a systematic review. *Front Oncol*. 2024;14:1343627. doi:10.3389/fonc.2024.1343627
104. Bobolea SS, Hinoveanu MI, Dimitriu A, et al. Exploring Artificial Intelligence's Potential to Enhance Conventional Anticancer Drug Development. *Drug Dev Res*. 2025;86(7):e70182. doi:10.1002/ddr.70182
105. Jiang B, Bao L, He S, Chen X, Jin Z, Ye Y. Deep learning applications in breast cancer histopathological imaging: diagnosis, treatment, and prognosis. *Breast Cancer Res*. 2024;26(1):137. doi:10.1186/s13058-024-01895-6
106. Liu C, Su J, Yang F, Wei K, Ma J, Zhou X. Compound signature detection on LINCS L1000 big data. *Mol Biosyst*. 2015;11(3):714–722. doi:10.1039/C4MB00677A
107. Firoozbakht F, Rezaeian I, Rueda L, Ngom A. Computationally repurposing drugs for breast cancer subtypes using a network-based approach. *BMC Bioinf*. 2022;23(1):143. doi:10.1186/s12859-022-04662-6
108. Karampuri A, Kundur S, Perugu S. Exploratory drug discovery in breast cancer patients: a multimodal deep learning approach to identify novel drug candidates targeting RTK signaling. *Comput Biol Med*. 2024;174:108433. doi:10.1016/j.compbiomed.2024.108433
109. Rezaee P, Rezaee S, Maaza M, Arab SS. Screening of BindingDB database ligands against EGFR, HER2, Estrogen, Progesterone and NF-κB receptors based on machine learning and molecular docking. *Comput Biol Med*. 2024;183:109279. doi:10.1016/j.compbiomed.2024.109279
110. Hamed AA, Fandy TE, Wu X, editors. Accelerating complex disease treatment through network medicine and GenAI: a case study on drug repurposing for breast cancer. 2024 IEEE International Conference on Medical Artificial Intelligence (MedAI); 2024: IEEE.
111. Podplutova E, Vepreva A, Konovalova OA, Vinogradov V, Shkil DO, Dmitrenko A. Pharmacophore-Guided Generative Design of Novel Drug-Like Molecules. arXiv preprint arXiv:251001480. 2025.
112. Wang ZF, Yang J, Di Li J, Lu SP. *AI-Driven Drug Discovery for ERBB2 Identified Through Network Pharmacology and GEO Data Analysis*. 2025.
113. Ouyang B, Shan C, Shen S, et al. AI-powered omics-based drug pair discovery for pyroptosis therapy targeting triple-negative breast cancer. *Nat Commun*. 2024;15(1):7560. doi:10.1038/s41467-024-51980-9
114. Zhou K, Tian B, Lu J, Dong B, Xu H. Machine learning-guided synthesis of nanomaterials for breast cancer therapy. *Sci Rep*. 2024;14(1):25795. doi:10.1038/s41598-024-76924-7
115. Dezsó Z, Ceccarelli M. Machine learning prediction of oncology drug targets based on protein and network properties. *BMC Bioinf*. 2020;21(1):104. doi:10.1186/s12859-020-3442-9
116. Al-Taie Z, Hannink M, Mitchem J, Papageorgiou C, Shyu C-R. Drug repositioning and subgroup discovery for precision medicine implementation in triple negative breast cancer. *Cancers*. 2021;13(24):6278. doi:10.3390/cancers13246278
117. Mohapatra R, Panda P, Rout SS. QSAR Based Drug Repurposing: a New Paradigm in Breast Cancer Research. *Ind J Pharm Edu Res*. 2024;58(3):695–708. doi:10.5530/ijper.58.3.78
118. Clare SE, Shaw PL. “Big Data” for breast cancer: where to look and what you will find. *NPJ Breast Cancer*. 2016;2(1):1–5. doi:10.1038/njbcancer.2016.31
119. Curtis C, Shah SP, Chin S-F, et al. The genomic and transcriptomic architecture of 2,000 breast tumours reveals novel subgroups. *Nature*. 2012;486(7403):346–352. doi:10.1038/nature10983
120. Gaulton A, Bellis LJ, Bento AP, et al. ChEMBL: a large-scale bioactivity database for drug discovery. *Nucleic Acids Res*. 2012;40(D1):D1100–D1107. doi:10.1093/nar/gkr777
121. Zdrzil B. Fifteen years of ChEMBL and its role in cheminformatics and drug discovery. *J Cheminf*. 2025;17(1):1–9. doi:10.1186/s13321-025-00963-z
122. Irwin JJ, Shoichet BK. ZINC— a free database of commercially available compounds for virtual screening. *J Chem Inf Model*. 2005;45(1):177–182.
123. Tingle BI, Tang KG, Castanon M, et al. ZINC-22— A free multi-billion-scale database of tangible compounds for ligand discovery. *J Chem Inf Model*. 2023;63(4):1166–1176. doi:10.1021/acs.jcim.2c01253
124. Clough E, Barrett T. The gene expression omnibus database. In: *Statistical Genomics: Methods and Protocols*. Springer; 2016:93–110.
125. Duan Q. *Harnessing the LINCS L1000 Data for Drug Discovery by Computational Analyses: Icahn School of Medicine at Mount Sinai*. 2016.
126. Zarin DA, Tse T, Williams RJ, Califf RM, Ide NC. The ClinicalTrials.gov results database—update and key issues. *N Engl J Med*. 2011;364(9):852–860. doi:10.1056/NEJMsa1012065
127. Consortium APG, André F, Arnedos M, Baras AS, et al. AACR Project GENIE: powering precision medicine through an international consortium. *Cancer Discovery*. 2017;7(8):818–831. doi:10.1158/2159-8290.CD-17-0151.
128. Kuno M, Osumi H, Udagawa S, et al. Artificial Intelligence in Clinical Oncology: from Productivity Enhancement to Creative Discovery. *Current Oncol*. 2025;32(11):588. doi:10.3390/curroncol32110588
129. Nemade V, Fegade V. Machine learning techniques for breast cancer prediction. *Procedia Comput Sci*. 2023;218:1314–1320. doi:10.1016/j.procs.2023.01.110
130. Xia Y, Sun M, Huang H, Jin W-L. Drug repurposing for cancer therapy. *Signal Transduc Target Ther*. 2024;9(1):92. doi:10.1038/s41392-024-01808-1
131. Zhavoronkov A, Ivanenkov YA, Aliper A, et al. Deep learning enables rapid identification of potent DDR1 kinase inhibitors. *Nature Biotechnol*. 2019;37(9):1038–1040. doi:10.1038/s41587-019-0224-x
132. Shahid MS, Imran A. Breast cancer detection using deep learning techniques: challenges and future directions. *Multimedia Tools Appl*. 2025;84(6):3257–3304. doi:10.1007/s11042-025-20606-7
133. Salehin I, Kang D-K. A review on dropout regularization approaches for deep neural networks within the scholarly domain. *Electronics*. 2023;12(14):3106. doi:10.3390/electronics12143106

134. Tiwari PC, Pal R, Chaudhary MJ, Nath R. Artificial intelligence revolutionizing drug development: exploring opportunities and challenges. *Drug Dev Res.* 2023;84(8):1652–1663. doi:10.1002/ddr.22115
135. Arrowsmith J. Phase II failures: 2008–2010. *Nat Rev Drug Discov.* 2011;10(5):328–329. doi:10.1038/nrd3439
136. DiMasi JA, Reichert JM, Feldman L, Malins A. Clinical approval success rates for investigational cancer drugs. *Clin Pharmacol Ther.* 2013;94(3):329–335. doi:10.1038/clpt.2013.117
137. Dwivedi R, Dave D, Naik H, et al. Explainable AI (XAI): core ideas, techniques, and solutions. *ACM Comput Surv.* 2023;55(9):1–33. doi:10.1145/3561048
138. Gunning D, Stefik M, Choi J, Miller T, Stumpf S, Yang G-Z. XAI—Explainable artificial intelligence. *Sci Rob.* 2019;4(37):eaay7120. doi:10.1126/scirobotics.aay7120
139. Biamonte J, Wittek P, Pancotti N, Rebentrost P, Wiebe N, Lloyd S. Quantum machine learning. *Nature.* 2017;549(7671):195–202. doi:10.1038/nature23474
140. Schuld M, Sinayskiy I, Petruccione F. An introduction to quantum machine learning. *Contemp Phys.* 2015;56(2):172–185. doi:10.1080/00107514.2014.964942
141. Choudhuri S, Yendluri M, Poddar S, et al. Recent advancements in computational drug design algorithms through machine learning and optimization. *Kinases and Phosphatases.* 2023;1(2):117–140. doi:10.3390/kinasesphosphatases1020008
142. Popova M, Isayev O, Tropsha A. Deep reinforcement learning for de novo drug design. *Sci Adv.* 2018;4(7):eaap7885. doi:10.1126/sciadv.aap7885
143. Wang H, Arulraj T, Ippolito A, Popel AS. From virtual patients to digital twins in immuno-oncology: lessons learned from mechanistic quantitative systems pharmacology modeling. *Npj Digital Med.* 2024;7(1):189. doi:10.1038/s41746-024-01188-4
144. Wu C, Lima EA, Stowers CE, et al. MRI-based digital twins to improve treatment response of breast cancer by optimizing neoadjuvant chemotherapy regimens. *Npj Digital Med.* 2025;8(1):195. doi:10.1038/s41746-025-01579-1
145. Moztarzadeh O, Jamshidi M, Sargolzaei S, et al. Metaverse and healthcare: machine learning-enabled digital twins of cancer. *Bioengineering.* 2023;10(4):455. doi:10.3390/bioengineering10040455
146. Khoshfekar Rudsari H, Tseng B, Zhu H, et al. Digital twins in healthcare: a comprehensive review and future directions. *Frontiers in Digital Health.* 2025;7:1633539. doi:10.3389/fgth.2025.1633539

Breast Cancer: Targets and Therapy

Publish your work in this journal

Breast Cancer - Targets and Therapy is an international, peer-reviewed open access journal focusing on breast cancer research, identification of therapeutic targets and the optimal use of preventative and integrated treatment interventions to achieve improved outcomes, enhanced survival and quality of life for the cancer patient. The manuscript management system is completely online and includes a very quick and fair peer-review system, which is all easy to use. Visit <http://www.dovepress.com/testimonials.php> to read real quotes from published authors.

Submit your manuscript here: <https://www.dovepress.com/breast-cancer—targets-and-therapy-journal>

Dovepress
Taylor & Francis Group