Content available at: https://www.ipinnovative.com/open-access-journals

Journal of Education Technology in Health Sciences

Journal homepage: https://jeths.org/

Review Article

Artificial Intelligence in Modern Deep-Learning QSAR

Shoheb Shaikh[1]*, Ashish Alte[1], Vishal Pande[1], Vaibhav Wagh[1], Jayprakash Suryawanshi[1]

[1]RSM NN Sattha College of Pharmacy Ahilyanagar, Maharashtra, India

Abstract

The integration of artificial intelligence (AI) with quantitative structure–activity relationship (QSAR) modeling has transformed drug discovery and medicinal chemistry by enabling rapid, accurate, and large-scale prediction of molecular properties and bioactivities. Traditional QSAR approaches are limited by linear assumptions, small datasets, and limited feature extraction capabilities. In contrast, modern deep-learning (DL) frameworks such as convolutional neural networks, graph neural networks, and recurrent architectures can automatically extract high-level chemical descriptors, handle complex nonlinearities, and generalize across diverse chemical spaces. This article critically reviews the evolution of QSAR from classical statistical models to AI-driven deep-learning paradigms, highlighting advances in molecular representation (e.g., SMILES embeddings, molecular graphs), model interpretability, and integration with cheminformatics workflows. We discuss recent applications of deep-learning QSAR in hit identification, lead optimization, ADMET prediction, and polypharmacology. Challenges such as data scarcity, overfitting, and the need for transparent, regulatory-compliant models are addressed. Finally, future directions are proposed for developing explainable and transferable AI-QSAR models capable of accelerating drug discovery in an era of big data and precision medicine.

Keywords: Artificial intelligence, Deep learning, Quantitative Structure–activity Relationship (QSAR), Drug discovery, Cheminformatics, Predictive modeling, Medicinal chemistry

Received: 12-10-2025; **Accepted:** 20-11-2025; **Available Online:** 27-11-2025

1. Introduction to QSAR and the Need for AI

1.1 Basics of QSAR

Quantitative Structure–Activity Relationship (QSAR) is a cornerstone of computational drug design, based on the principle that the biological activity of a compound is a function of its chemical structure. QSAR methods attempt to establish mathematical models that correlate molecular descriptors such as physicochemical, topological, electronic, or hydrophobic parameters with pharmacological activities or toxicological profiles of compounds.[1]

The QSAR approach has been widely applied in drug discovery, toxicology, agrochemical research, and environmental sciences. For instance, hydrophobicity (expressed as logP), electronic distribution (Hammett constants), and steric effects (molar refractivity) have been employed as classical descriptors in building predictive models.[2] Techniques such as linear regression, partial least squares (PLS), and principal component analysis (PCA) are traditionally used for generating QSAR equations.[3] The regulatory acceptance of QSAR by bodies like the U.S. Environmental Protection Agency (EPA) and the Organisation for Economic Co-operation and Development (OECD) further underscores its importance in chemical risk assessment.[4]

By providing a cost-effective and time-saving alternative to extensive in vitro and in vivo experimentation, QSAR has played a significant role in early-stage virtual screening and lead optimization.[5]

1.2. Limitations of classical QSAR

Despite its success, classical QSAR suffers from several limitations. First, the selection of descriptors is often subjective and restricted to pre-defined chemical properties, which may not fully capture the underlying complexity of molecular

*Corresponding author: Shoheb Shaikh
Email: research.publication8891@gmail.com

https://doi.org/10.18231/j.jeths.30178.1764230559

interactions.[6] Linear and semi-linear models frequently fail to accommodate the non-linear, high-dimensional relationships between structure and activity. Moreover, the extrapolation ability of traditional QSAR is weak, as models built on specific chemical series often lack generalizability when applied to structurally diverse datasets.[7]

Another major challenge is the curse of dimensionality. With the availability of thousands of molecular descriptors, feature redundancy and noise can lead to overfitting, resulting in poor predictive accuracy for external datasets. Additionally, small datasets a common scenario in rare disease drug discovery limit the statistical robustness of classical QSAR approaches. **(Figure 1)**

Interpretability is also a concern. While simple QSAR equations are relatively transparent, complex multi-parameter models can be difficult to interpret mechanistically, reducing their usefulness in guiding medicinal chemistry decisions.[8] Furthermore, classical QSAR methods have limited ability to capture dynamic biological processes such as protein conformational changes, solvation effects, or allosteric modulations.[9]

These constraints have motivated researchers to explore more flexible, data-driven methods capable of handling large, heterogeneous datasets and uncovering non-linear relationships that classical QSAR cannot.

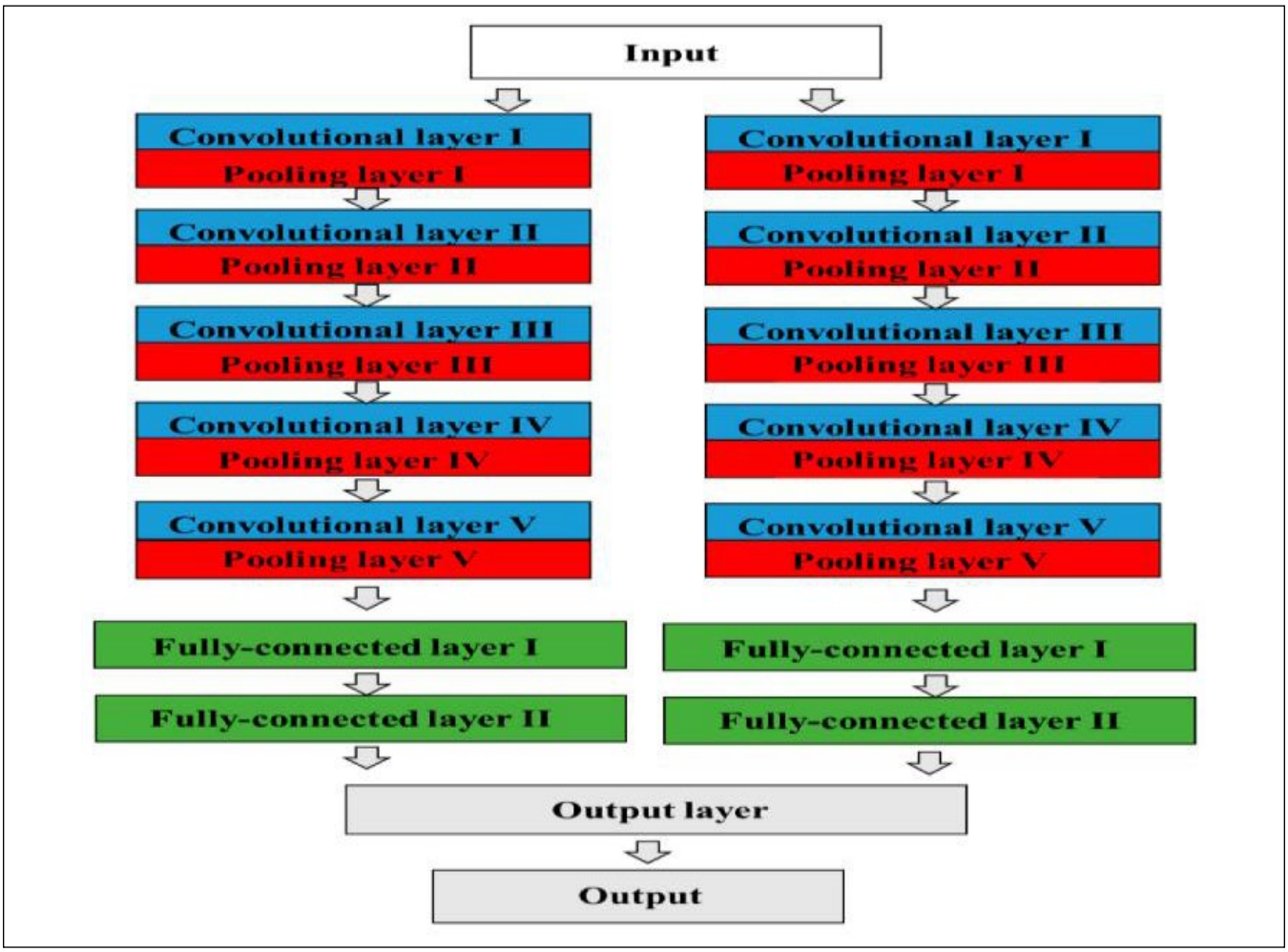

Figure 1: Classical QSAR vs. Deep-learning–enhanced QSAR: A schematic comparing traditional QSAR methods (Linear regression, descriptor-based) with modern AI/deep-learning approaches (CNNs, RNNs, GNNs).

1.3. Rise of AI and deep learning in cheminformatics

The recent surge of Artificial Intelligence (AI) and deep learning has transformed cheminformatics by overcoming many of the limitations inherent in traditional QSAR. AI-driven models such as random forests, support vector machines, and ensemble learning techniques have demonstrated superior performance in handling high-dimensional chemical data, enabling accurate predictions across diverse compound libraries.[10]

Deep learning, particularly neural networks with multiple hidden layers, has further advanced predictive modeling by learning complex, non-linear relationships without requiring explicit feature engineering.[11] Architectures such as convolutional neural networks (CNNs) and graph neural networks (GNNs) can directly process molecular graphs or SMILES strings, thereby eliminating the need for manually curated descriptors.[12] This capability allows models to capture intrinsic structural features, chemical substructures, and molecular interactions with unprecedented precision.

AI integration in QSAR has also enabled multi-task learning, transfer learning, and active learning strategies that are particularly useful for drug repurposing, de novo drug design, and addressing sparse data challenges in rare diseases.[13] Additionally, explainable AI approaches are being developed to enhance model interpretability, bridging the gap between predictive accuracy and mechanistic understanding.[14]

The rise of AI in QSAR represents not just an incremental improvement but a paradigm shift. By leveraging big data, computational power, and sophisticated algorithms, AI-driven QSAR models hold the promise of accelerating drug discovery pipelines, reducing experimental costs, and supporting personalized medicine initiatives.[15]

2. Fundamentals of Deep Learning for QSAR

2.1 Overview of neural networks

Artificial Neural Networks (ANNs) are the foundational models of deep learning, designed to emulate the way biological neurons process and transmit information. In their simplest form, ANNs consist of layers of interconnected units called artificial neurons or nodes. Each node performs a weighted summation of inputs, followed by the application of a non-linear activation function, producing an output that is propagated to subsequent layers.[16]

The basic architecture of a neural network comprises three key components:

1. **Input layer:** Tepresenting the molecular descriptors or features (e.g., physicochemical, topological, or graph-based features of chemical compounds).
2. **Hidden layers:** Where computations are performed through successive transformations, enabling the model to capture complex, non-linear relationships between chemical structure and biological activity.
3. **Output layer:** Which provides the final prediction, such as biological activity, toxicity, or binding affinity.[17]

In the context of QSAR, the power of neural networks lies in their ability to approximate highly non-linear functions. Unlike classical QSAR methods that rely on predefined descriptors and linear models, ANNs can uncover latent interactions between variables that may not be evident through conventional statistical approaches.[18] For instance, non-linear activation functions such as sigmoid, hyperbolic tangent (tanh), or rectified linear unit (ReLU) enable networks to capture complex structure–activity landscapes more accurately.[19]

The depth of a network refers to the number of hidden layers. While shallow networks (single hidden layer) are capable of modeling simple relationships, deep neural networks (DNNs) with multiple hidden layers offer hierarchical feature extraction. In cheminformatics, this means that lower layers can capture fundamental molecular features (such as atom-level connectivity or substructures), while deeper layers can learn higher-order interactions relevant to pharmacological activity.[20]

Training a neural network involves optimizing the weights of connections between nodes using algorithms such as backpropagation, which relies on gradient descent to minimize the difference between predicted and actual outputs.[21] The availability of large chemical datasets, improved computational power through GPUs, and algorithmic advancements have contributed to the resurgence of neural networks in modern QSAR applications.[22]

A significant advantage of neural networks in QSAR is their flexibility. They can handle diverse forms of chemical data, from traditional molecular descriptors to raw representations such as Simplified Molecular Input Line Entry System (SMILES) strings or molecular graphs.[23] Moreover, neural networks can be extended into specialized architectures like convolutional neural networks (CNNs) for capturing spatial molecular features and graph neural networks (GNNs) for directly modeling molecular graphs.[24]

2.2. CNNs, RNNs, and GNNs in molecular modelling

Convolutional Neural Networks (CNNs), originally designed for image recognition, have been adapted to molecular modelling by treating chemical structures as two-dimensional matrices (e.g., molecular fingerprints or grids of atomic features). CNNs are effective at identifying local substructural patterns, enabling prediction of bioactivity and toxicity.[25]

Recurrent Neural Networks (RNNs) are well suited for sequential data, making them particularly effective in processing SMILES strings, where each character represents atoms or bonds. Variants such as Long Short-Term Memory (LSTM) networks address long-range dependencies, enabling accurate modeling of complex molecular sequences.[26]

Graph Neural Networks (GNNs) represent a significant advancement, as molecules can be naturally encoded as graphs with atoms as nodes and bonds as edges. GNNs learn directly from molecular topology, capturing both local and global structural properties. Techniques such as message passing neural networks (MPNNs) and graph convolutional networks (GCNs) have achieved state-of-the-art performance in molecular property prediction and QSAR tasks.[27]

Collectively, CNNs, RNNs, and GNNs provide complementary strengths in molecular modelling, offering flexible frameworks to process diverse chemical data formats and improving predictive accuracy beyond classical QSAR.

2.3. Autoencoders and GANs for molecular design

Autoencoders are unsupervised neural architectures that learn compressed representations of molecules by encoding high-dimensional chemical data into a latent space and reconstructing it back. In QSAR and drug discovery, autoencoders enable dimensionality reduction, denoising of molecular descriptors, and generation of novel compound

structures from latent features.[28] Variational Autoencoders (VAEs) extend this framework by learning continuous latent distributions, facilitating smooth exploration of chemical space for de novo molecular design.[29]

Generative Adversarial Networks (GANs) consist of a generator that proposes novel molecular structures and a discriminator that evaluates their plausibility. In cheminformatics, GANs have been applied to design drug-like molecules with optimized properties such as solubility, bioavailability, and activity against specific targets.[30] The adversarial training process encourages the generator to produce chemically valid and diverse structures, accelerating lead discovery.[31] **(Figure 2)**

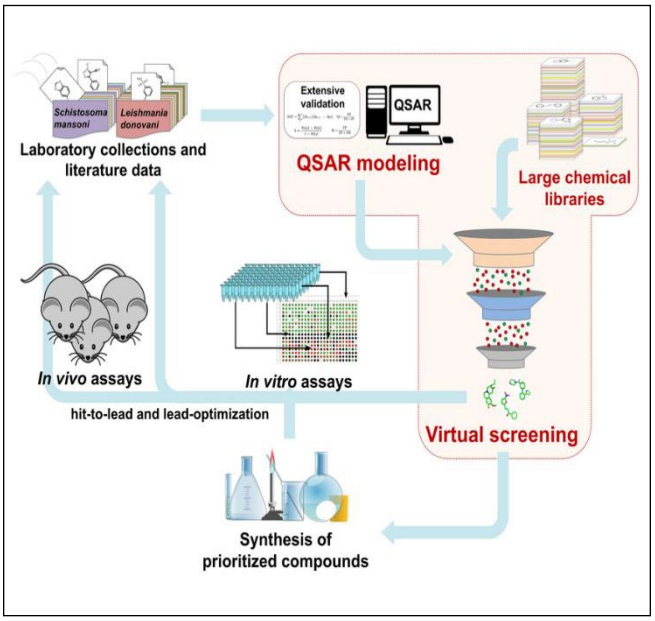

Figure 2: General workflow of AI-Driven QSAR modeling: Data collection → preprocessing → feature extraction → model building → validation → prediction.

3. Data Representation and Descriptor Engineering

The predictive success of QSAR and deep learning models strongly depends on how chemical structures are represented and encoded for computational analysis. Data representation serves as the bridge between molecular structures and machine learning algorithms, influencing both model performance and interpretability.

Classical QSAR traditionally relies on molecular descriptors, which are numerical values derived from chemical structures that capture physicochemical, topological, electronic, and geometric properties.[32] Examples include logP for lipophilicity, molecular weight, polar surface area, and descriptors derived from quantum chemistry calculations. More advanced representations include 2D topological indices, 3D conformational descriptors, and 4D/5D descriptors that account for dynamic molecular states.[33]

In cheminformatics, molecular fingerprints are widely used as binary or count-based representations of substructures within molecules. Popular variants such as MACCS keys, Extended Connectivity Fingerprints (ECFPs), and Morgan fingerprints encode substructural patterns for similarity search and QSAR modeling.[34] These fingerprints provide compact, high-dimensional representations that are computationally efficient and interpretable in terms of chemical subunits.

With the rise of deep learning, representation learning has shifted toward using raw formats like Simplified Molecular Input Line Entry System (SMILES) strings or molecular graphs. SMILES enable sequence-based modeling using recurrent neural networks (RNNs) or transformers,[35] while graph-based encodings allow graph neural networks (GNNs) to directly exploit atomic connectivity and bond features.[27] These approaches bypass handcrafted descriptors by learning task-specific features directly from molecular structure.

Descriptor engineering remains crucial, as redundancy, correlation, and noise in descriptor sets can impair model accuracy. Feature selection and dimensionality reduction techniques such as principal component analysis (PCA) and autoencoders are often applied to refine descriptor sets for robust QSAR modeling.[36]

3.1. Molecular fingerprints and smiles encoding

Molecular fingerprints are one of the most widely used chemical representations in QSAR modeling and cheminformatics. They encode the presence or absence of specific structural fragments as binary vectors, or in some cases as count-based vectors, thereby enabling efficient similarity searches and predictive modeling.[37] Popular fingerprints include MACCS keys, which use a fixed set of predefined structural motifs, and Extended Connectivity Fingerprints (ECFPs), also known as circular fingerprints, which capture atom-centered substructures within a defined radius.[38] These representations are particularly effective for substructure analysis, clustering, and virtual screening.

Complementary to fingerprints, the Simplified Molecular Input Line Entry System (SMILES) encodes molecules as one-dimensional strings based on rules for atoms, bonds, and connectivity.[39] SMILES strings are compact and easily parsed, making them compatible with sequence-based learning models such as recurrent neural networks (RNNs) and transformers. However, SMILES are not unique (multiple strings may represent the same molecule), which has led to the development of canonical SMILES and randomized SMILES strategies for improved model generalization.[40]

Both fingerprints and SMILES representations have been extensively applied in QSAR, offering flexible and computationally efficient means of capturing molecular information. While fingerprints are well-suited for similarity-driven tasks, SMILES encoding facilitates deep learning applications, particularly in generative molecular design. **(Figure 3)**

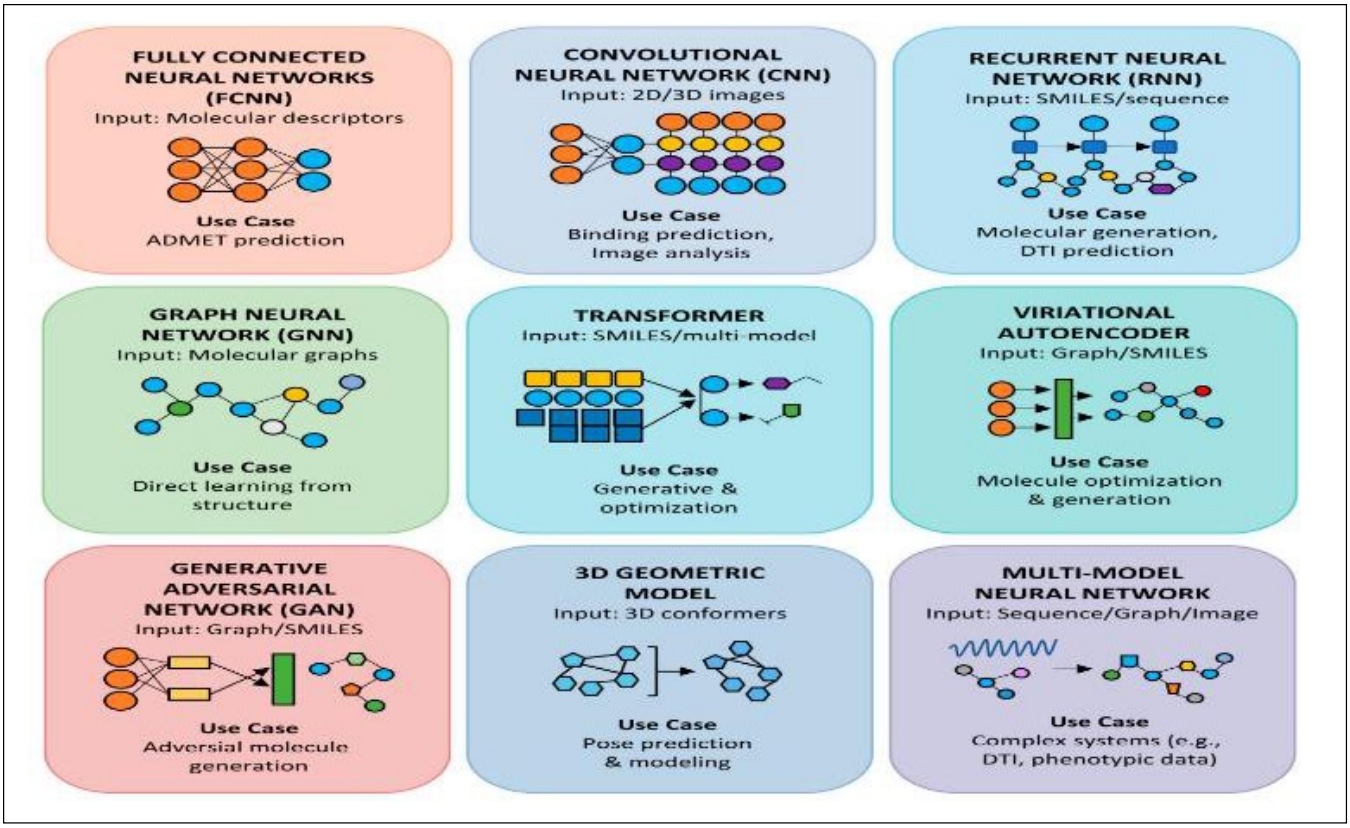

Figure 3: Types of deep learning architectures in QSAR: CNN, RNN, autoencoders, graph neural networks — with examples of molecular input representations (SMILES, fingerprints, molecular graphs).

3.2. Feature extraction using deep learning

Deep learning introduces a paradigm shift in molecular feature representation by enabling automatic feature extraction directly from raw molecular data, reducing reliance on handcrafted descriptors. Instead of predefining structural or physicochemical parameters, neural networks learn hierarchical, task-specific features that capture both local and global chemical patterns.

For sequence-based inputs like SMILES, convolutional neural networks (CNNs) and recurrent neural networks (RNNs) can automatically extract chemical substructures, bond contexts, and sequential dependencies relevant for bioactivity prediction. For molecular graphs, Graph Neural Networks (GNNs) apply message-passing algorithms to propagate information between atoms and bonds, learning expressive node and graph-level embeddings.

Unsupervised architectures such as autoencoders also contribute to feature extraction by compressing high-dimensional molecular descriptors into compact latent spaces, which preserve essential information while filtering noise. Similarly, attention-based models, including transformers, enhance interpretability by highlighting the most informative molecular substructures associated with activity.[41]

3.3. Graph-Based molecular representations

Molecules can be naturally represented as graphs, where atoms correspond to nodes and chemical bonds to edges. This representation captures both local and global structural properties of molecules more explicitly than traditional descriptors or SMILES strings. Each node is associated with features such as atomic number, hybridization state, valence, and charge, while edges can encode bond type, order, and aromaticity.[42] Unlike linear encodings, graph-based representations inherently preserve molecular topology and connectivity, making them highly expressive for QSAR and cheminformatics tasks.

The rise of Graph Neural Networks (GNNs) has enabled direct learning from molecular graphs without the need for handcrafted descriptors. GNNs operate through message passing, where information is iteratively exchanged between neighboring atoms and aggregated to generate node-level and graph-level embeddings. Variants such as Graph Convolutional Networks (GCNs), Graph Attention Networks (GATs), and Message Passing Neural Networks (MPNNs) have shown state-of-the-art performance in predicting physicochemical properties, bioactivity, and toxicity.[43]

Graph-based approaches also facilitate the modeling of context-dependent chemical interactions, including stereochemistry and ring structures, which are challenging for traditional fingerprint methods. Additionally, GNNs can incorporate 3D spatial information, enabling the prediction of quantum chemical properties and protein–ligand interactions.[44] Recent extensions such as equivariant GNNs integrate geometric constraints, making them particularly relevant for molecular dynamics and structure-based drug design.[45]

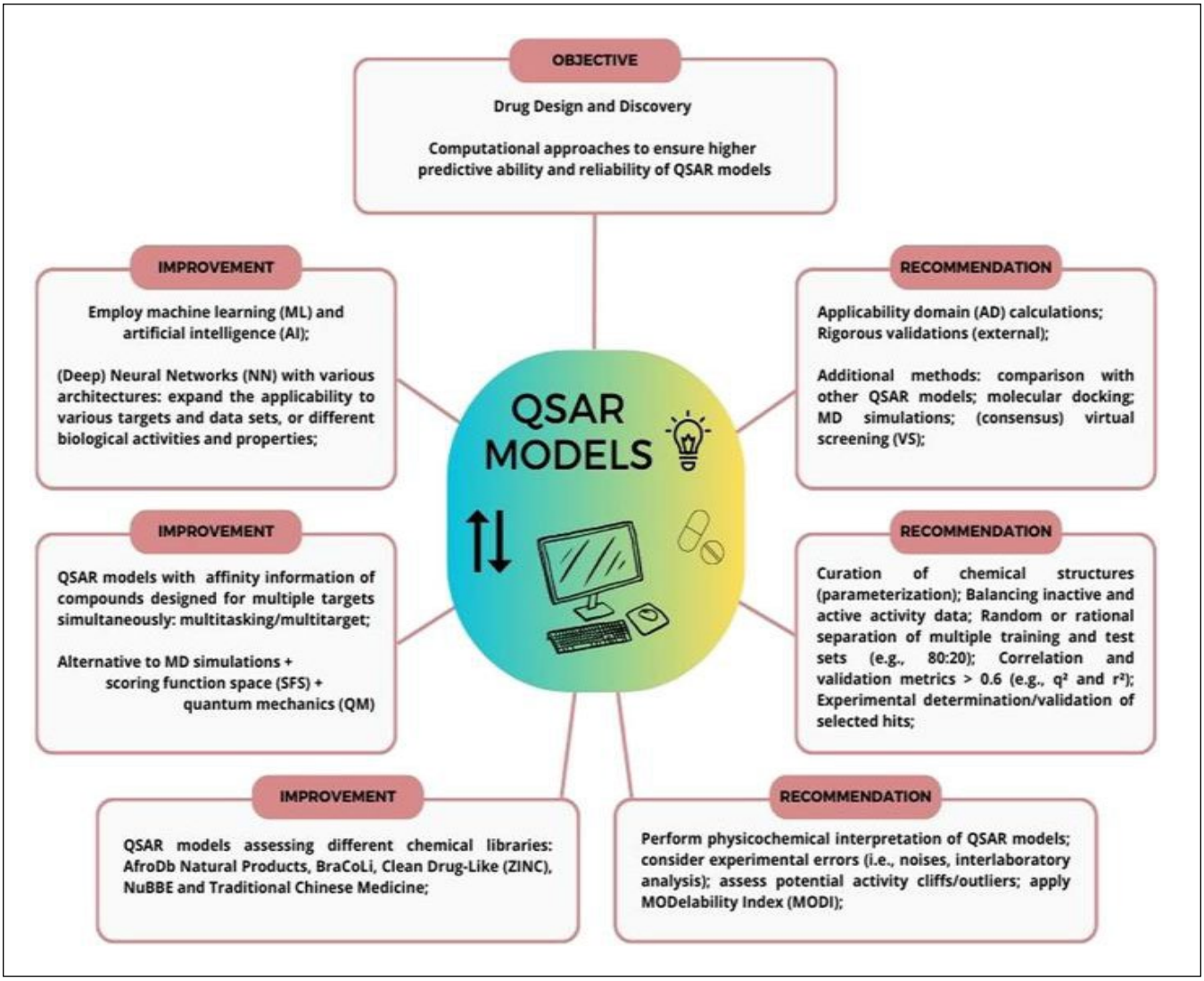

Figure 4: Application of deep QSAR in drug discovery: Case studies or schematic of virtual screening, ADMET prediction, toxicity assessment, and hit-to-lead optimization

4. AI-Powered QSAR Modeling Frameworks

The integration of Artificial Intelligence (AI) into QSAR modeling has revolutionized how chemical and biological data are processed, analyzed, and utilized in drug discovery. Unlike traditional QSAR approaches, which rely heavily on handcrafted descriptors and linear modeling, AI-powered frameworks employ machine learning (ML) and deep learning algorithms to capture highly non-linear and multidimensional relationships between molecular structures and biological activities. A typical AI-powered QSAR framework consists of four core components.

4.1. Data preprocessing and representation

Molecules are encoded using descriptors, fingerprints, SMILES strings, or graph-based representations.[46] Model construction – AI algorithms such as random forests, support vector machines, convolutional neural networks (CNNs), recurrent neural networks (RNNs), or graph neural networks (GNNs) are trained on chemical datasets to learn structure–activity relationships. Validation and evaluation – frameworks employ cross-validation, external validation sets, and applicability domain analysis to ensure robustness and generalizability. Deployment and interpretation – predictive models are integrated into pipelines for virtual screening, lead optimization, and toxicity prediction, with increasing emphasis on explainable AI for interpretability.

Several specialized frameworks have been developed to support AI-driven QSAR. DeepChem provides an open-source platform integrating GNNs and multitask learning for molecular property prediction. MoleculeNet, a benchmark dataset suite, facilitates standardized evaluation of deep learning models for QSAR.[47] Commercial platforms such as BIOVIA, Schrodinger's DeepChemistry, and Atomwise leverage proprietary AI pipelines for large-scale screening and de novo design.

The advantages of these frameworks include scalability to millions of compounds, improved predictive accuracy, and the ability to handle multimodal data, such as integrating chemical structures with biological assay results or omics data. However, challenges remain in terms of interpretability,

reproducibility, and regulatory acceptance. Efforts toward explainable AI and transfer learning are addressing these gaps, making AI-powered QSAR increasingly relevant for precision medicine and rational drug design.[48]

4.2. End-to-end deep learning models

End-to-end deep learning has emerged as a powerful paradigm in QSAR, enabling models to directly learn predictive features from raw chemical representations without manual descriptor engineering. Instead of relying on predefined physicochemical or structural descriptors, these models process inputs such as SMILES strings or molecular graphs, automatically extracting hierarchical features relevant for activity prediction.[49]

Convolutional neural networks (CNNs) and recurrent neural networks (RNNs) have been applied to SMILES-based representations, learning substructural patterns and long-range dependencies between atoms and bonds. More recently, graph neural networks (GNNs) have gained prominence by leveraging molecular graphs, where message-passing algorithms allow direct learning of atomic and bond-level interactions.

End-to-end models offer significant advantages in scalability and generalizability, as they can adapt to diverse chemical datasets and capture complex non-linear structure–activity relationships. However, challenges remain in terms of interpretability, data requirements, and potential overfitting when applied to small or imbalanced datasets. **(Figure 4)**

4.3. Transfer learning and multi-task learning in QSAR

Transfer learning and multi-task learning (MTL) are strategies that enhance QSAR modeling by leveraging shared knowledge across datasets and prediction tasks. Transfer learning involves pretraining a model on a large, chemically diverse dataset, followed by fine-tuning on a smaller, task-specific dataset.[50] This approach is particularly valuable in rare disease drug discovery, where limited labeled data hampers traditional QSAR modeling.

Multi-task learning enables simultaneous prediction of multiple molecular properties (e.g., solubility, toxicity, and activity), with tasks sharing common representations across neural network layers.[51] MTL improves generalization by exploiting correlations between tasks and mitigating overfitting, especially when individual datasets are small.

Applications of transfer and MTL approaches include large-scale bioactivity prediction, ADMET modeling, and drug repurposing. Frameworks such as DeepChem and MoleculeNet benchmarks have demonstrated the effectiveness of these strategies in improving both predictive performance and computational efficiency.

4.4. Software and platforms (DeepChem, Chemprop)

The rapid adoption of AI in QSAR has been facilitated by the development of open-source platforms and specialized toolkits that lower barriers to entry and standardize workflows. DeepChem is one of the most widely used Python-based frameworks for molecular machine learning. It provides a unified interface for diverse algorithms, including graph convolutional networks, multitask learning, and transfer learning, alongside benchmark datasets from MoleculeNet.[52] DeepChem supports end-to-end pipelines encompassing molecular featurization, model training, evaluation, and deployment, making it a valuable resource for academic and industrial research.

Chemprop is another state-of-the-art platform designed for molecular property prediction using message-passing neural networks (MPNNs).[53] It employs directed MPNNs for feature extraction from molecular graphs, achieving strong performance across bioactivity and physicochemical property prediction benchmarks. Chemprop supports multi-task learning, uncertainty quantification, and ensemble methods, providing robust and scalable QSAR modeling capabilities.[54]

5. Model Training, Validation, and Evaluation

The reliability of AI-powered QSAR models depends not only on algorithmic design but also on rigorous training, validation, and evaluation strategies. During training, molecular representations (e.g., fingerprints, SMILES, graphs) are mapped to biological or physicochemical activities through supervised learning. Hyperparameters such as learning rate, batch size, and network depth are optimized to balance predictive accuracy and generalization.

Validation is crucial to avoid overfitting and to assess the model's ability to generalize. Common strategies include k-fold cross-validation, leave-one-out validation, and the use of external test sets that remain unseen during training. The concept of an applicability domain is also vital, defining the chemical space in which the model's predictions are reliable.[55]

For evaluation, multiple performance metrics are used depending on the task type. Regression tasks employ root mean square error (RMSE), mean absolute error (MAE), and coefficient of determination (R^2), while classification tasks rely on accuracy, area under the ROC curve (AUC), precision, recall, and F1-score.[56] More advanced metrics, such as calibration curves and uncertainty quantification, are increasingly integrated to improve decision-making in drug discovery pipelines.

5.1. Training pipelines and hyperparameter tuning

The development of AI-powered QSAR models relies on well-designed training pipelines that ensure data quality, model robustness, and reproducibility. A typical pipeline begins with data preprocessing, involving curation, removal of duplicates, normalization, and molecular encoding into suitable formats such as fingerprints, SMILES strings, or graph-based features. Following this, the dataset is divided into training, validation, and test sets to prevent information leakage and ensure unbiased performance estimation.

Model training involves selecting an appropriate neural architecture (e.g., convolutional, recurrent, or graph neural networks) and optimizing it on the training data. Central

to this process is hyperparameter tuning, which controls aspects such as learning rate, batch size, number of layers, hidden units, dropout probability, and optimizer choice.[57] Hyperparameter values strongly influence convergence, generalization, and computational efficiency.

Conventional tuning methods include grid search and random search, which systematically explore parameter spaces but may be computationally expensive. More advanced methods, such as Bayesian optimization, genetic algorithms, and Hyperband, improve efficiency by adaptively exploring promising regions of the search space.[58] Recently, AutoML frameworks have automated this process, accelerating the design of optimized QSAR pipelines.[59] **(Figure 5)**

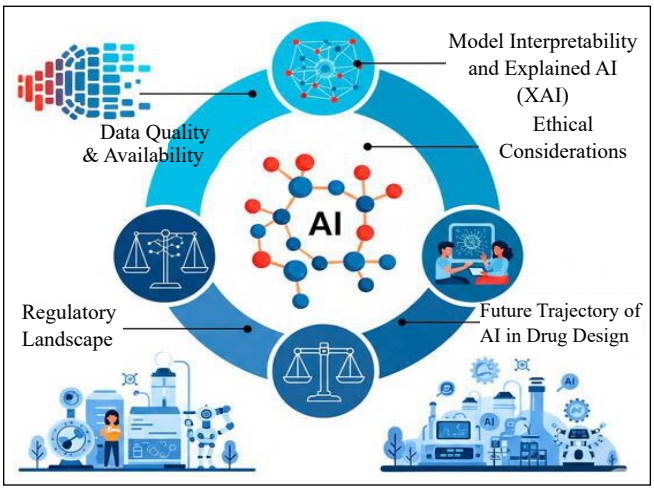

Figure 5: Future Perspective: Integration of QSAR with AI tools (federated learning, explainable AI, quantum computing).

5.2. Cross-validation and external validation techniques

Validation is critical in QSAR modeling to ensure that predictive performance is not overestimated and that models generalize to unseen compounds. Cross-validation (CV) is the most widely used internal validation strategy. In k-fold CV, the dataset is partitioned into *k* subsets; the model is trained on *k-1* folds and tested on the remaining fold, with the process repeated until each fold has served as the test set. Variants include leave-one-out CV (LOO-CV), useful for small datasets, and stratified CV, which preserves class balance in classification problems.[60]

However, cross-validation alone may not reflect real-world predictive reliability. External validation testing on a dataset completely excluded from model building provides a stronger measure of generalizability. A common strategy is the temporal split, where training is performed on earlier compounds and validation on more recent ones, mimicking prospective prediction in drug discovery.

5.3. Metrics for QSAR performance evaluation

Selecting the right evaluation metrics is essential for comparing and interpreting QSAR model performance. The choice depends on whether the task is regression (predicting continuous molecular properties) or classification (predicting categorical outcomes such as active/inactive compounds).

For regression tasks, metrics include Root Mean Square Error (RMSE), Mean Absolute Error (MAE), and R^2 (coefficient of determination). RMSE and MAE quantify the magnitude of prediction errors, while R^2 evaluates the proportion of variance explained by the model.[61]

For classification tasks, commonly used metrics are accuracy, precision, recall, and the F1-score, which balance sensitivity and specificity.[62] In imbalanced datasets, Area Under the ROC Curve (AUC-ROC) and Area Under the Precision-Recall Curve (AUC-PR) provide more informative measures of discriminative ability.[63]

Beyond conventional metrics, calibration plots, applicability domain analysis, and uncertainty estimation are increasingly applied to assess the reliability of predictions in drug discovery. These measures not only evaluate accuracy but also support regulatory acceptance of QSAR models.

6. Applications in Drug Discovery and Toxicology

AI-powered QSAR models are increasingly applied in drug discovery and toxicology, where they accelerate compound screening, optimize lead molecules, and ensure safety assessment. In early-stage drug discovery, QSAR models help predict bioactivity, binding affinity, and selectivity against therapeutic targets, thereby reducing reliance on costly and time-consuming high-throughput screening. Deep learning frameworks, such as convolutional and graph neural networks, enable accurate modeling of large chemical libraries, improving the identification of novel drug candidates with desirable pharmacological profiles.[64]

In toxicology, QSAR approaches are widely employed for in silico toxicity prediction, including carcinogenicity, hepatotoxicity, cardiotoxicity, and endocrine disruption. Regulatory agencies such as the U.S. Environmental Protection Agency (EPA) and the European Chemicals Agency (ECHA) endorse QSAR models as alternatives to animal testing under REACH and other safety guidelines. Moreover, multi-task learning allows simultaneous prediction of multiple toxicity endpoints, enhancing risk assessment efficiency.

6.1. Hit identification and lead optimization

In modern drug discovery, QSAR models are integral to hit identification and lead optimization, where they reduce experimental workload by prioritizing promising molecules. Deep learning–based QSAR frameworks can analyze millions of virtual compounds to identify hits with high binding affinity to specific biological targets. For example, convolutional neural networks (CNNs) trained on molecular fingerprints and SMILES representations have shown superior performance in virtual screening compared to classical approaches.

Once hits are identified, QSAR models further aid in lead optimization, predicting activity cliffs, structure–

activity relationships, and guiding chemical modifications to improve potency, selectivity, and safety. Graph neural networks (GNNs) are particularly valuable because they directly capture molecular topology, allowing optimization of structural features critical for drug–target interactions. Moreover, AI-driven multi-parameter optimization integrates QSAR with pharmacokinetics and toxicity endpoints, enabling rational decision-making in lead development.[65]

6.2. Toxicity prediction and admet profiling

QSAR models are increasingly deployed for toxicity prediction and ADMET (Absorption, Distribution, Metabolism, Excretion, and Toxicity) profiling, critical steps in ensuring drug safety. Traditional in vitro and in vivo assays are resource-intensive and ethically challenging; hence, in silico QSAR models serve as efficient, non-animal alternatives.[66]

Machine learning–based QSAR approaches predict diverse toxicity endpoints, including carcinogenicity, cardiotoxicity, hepatotoxicity, and endocrine disruption.[67] Multi-task neural networks have demonstrated the ability to predict multiple ADMET properties simultaneously, improving efficiency and reducing data fragmentation.[68] For example, deep learning frameworks trained on large pharmacokinetic datasets have outperformed conventional models in predicting blood–brain barrier permeability and cytochrome P450 interactions.

Regulatory agencies such as the U.S. Environmental Protection Agency (EPA) and European Chemicals Agency (ECHA) endorse QSAR-based toxicity prediction for chemical safety assessment, especially under REACH and OECD guidelines. Integrating AI-driven QSAR into ADMET profiling enables early elimination of unsafe candidates, reducing late-stage clinical failures.

6.3. Personalized QSAR models

The integration of personalized medicine with QSAR modeling is a growing frontier in precision drug discovery. Traditional QSAR models generalize across populations, but they often overlook inter-individual variability in drug response due to genetic, epigenetic, and metabolic differences .[69] Personalized QSAR models aim to incorporate patient-specific data, such as genomics, proteomics, and metabolomics, to predict drug efficacy and safety tailored to individual profiles.

AI-based QSAR approaches, especially those leveraging deep learning, enable the integration of multi-omics data with molecular descriptors, capturing patient-specific variability in drug response.[70] For example, machine learning models trained on cancer genomics and pharmacogenomics datasets (e.g., GDSC, CCLE) can predict how different patient-derived cell lines respond to targeted therapies.[71] This facilitates the design of personalized drug regimens and repurposing of existing drugs for specific subgroups.

Although still in early stages, personalized QSAR holds promise in reducing adverse effects, improving therapeutic outcomes, and supporting precision oncology and individualized toxicology testing.[72]

7. Challenges and Future Perspectives

Despite remarkable progress, the application of AI-driven QSAR modeling faces several challenges that limit its widespread adoption. A major concern is data quality and availability, as QSAR models rely heavily on large, diverse, and accurately annotated datasets. Many chemical and biological datasets are biased, imbalanced, or contain experimental inconsistencies, which can compromise model reliability. Another challenge lies in the interpretability of deep learning models. While neural networks provide high predictive accuracy, their “black-box” nature often hinders mechanistic insights, raising concerns for regulatory acceptance.

Model generalizability is another critical issue; QSAR models trained on limited chemical space often fail when extrapolated to novel compounds. Furthermore, hyperparameter optimization and reproducibility remain difficult, particularly when dealing with complex deep learning architectures. On the regulatory side, agencies demand transparency and explainability in predictive models for drug approval and chemical safety assessment, creating barriers for AI-driven QSAR integration.[73]

Looking forward, emerging trends such as transfer learning, federated learning, and explainable AI (XAI) hold promise for overcoming these barriers. Integration of QSAR with multi-omics, systems biology, and real-world patient data will further enhance personalized drug discovery. Collaboration between academia, industry, and regulatory bodies will be vital to ensure robust, interpretable, and clinically relevant AI-powered QSAR models.

7.1. Data quality and size limitations

The reliability of QSAR models depends strongly on the quality, diversity, and volume of chemical–biological datasets. Many public datasets contain noisy, imbalanced, or inconsistent measurements, limiting model robustness and reproducibility.[74] Furthermore, small dataset sizes remain a barrier for deep learning, which generally requires large-scale data to avoid overfitting. Initiatives such as ChEMBL and PubChem provide curated chemical–bioactivity data, but challenges remain in harmonizing assay protocols and addressing data sparsity for rare targets.[75] Federated learning and data-sharing frameworks are emerging to mitigate these limitations without compromising data privacy.

7.2 Model interpretability and regulatory acceptance

Although deep learning–based QSAR models achieve high predictive performance, their “black-box” nature raises concerns about trustworthiness in decision-making. For regulatory agencies such as the FDA, EMA, and OECD, interpretability is a key requirement for risk assessment and safety validation. Lack of transparency hinders the acceptance of AI-driven QSAR in regulatory submissions for drug

approval and toxicology. Explainable AI (XAI) methods, including attention mechanisms, saliency mapping, and Shapley additive explanations (SHAP), are being developed to provide mechanistic insights while retaining predictive accuracy. Bridging this gap is essential for translating QSAR models into regulatory practice.

7.3 Integration with generative AI and quantum computing

Recent advances in generative AI (e.g., variational autoencoders, GANs, and diffusion models) have enabled the design of novel molecules optimized for activity, safety, and synthesizability. When combined with QSAR models, generative frameworks create closed-loop drug discovery pipelines capable of de novo molecular design and optimization.[76] In parallel, quantum computing offers potential breakthroughs by simulating quantum–chemical interactions beyond the capacity of classical computing, thus improving descriptor generation and activity prediction.[77] Although still in its infancy, the integration of QSAR with quantum machine learning (QML) and generative AI is poised to redefine next-generation drug discover.

8. Conclusion

AI-powered deep-learning methods have enhanced QSAR modeling by enabling richer molecular representations and more accurate predictions, improving multiple stages of drug discovery. Continued advances in interpretability, data quality, and model robustness will be essential for fully realizing the potential of AI-QSAR in precision medicine.

9. Source of Funding

None.

10. Conflict of Interest

None.

References

1. Hansch C, Fujita T. ρ-σ-π Analysis. A method for the correlation of biological activity and chemical structure. *J Am Chem Soc*. 1964;86(8):1616–26. https://doi.org/10.1021/ja01062a035
2. Kubinyi H. QSAR and 3D QSAR in drug design part 1: Methodology. *Drug Discov Today*. 1997;2(11):457–67. http://dx.doi.org/10.1016/S1359-6446(97)01079-9
3. Eriksson L, Jaworska J, Worth AP, Cronin MTD, McDowell RM, Gramatica P. Methods for reliability and uncertainty assessment and for applicability evaluations of classification- and regression-based QSARs. *Environ Health Perspect*. 2003;111(10):1361–75. https://doi.org/10.1289/ehp.5758
4. OECD. Guidance document on the validation of (quantitative) structure–activity relationship [(Q)SAR] models. OECD Environment Health and Safety Publications, Series on Testing and Assessment. 2007;69:1–154. Available from: https://www.oecd.org/en/publications/guidance-document-on-the-validation-of-quantitative-structure-activity-relationship-q-sar-models_9789264085442-en.html
5. Cherkasov A, Muratov EN, Fourches D, Varnek A, Baskin II, Cronin M, et al. QSAR modeling: where have you been? Where are you going to? *J Med Chem*. 2014;57(12):4977–5010. https://doi.org/10.1021/jm4004285
6. Tropsha A. Best practices for QSAR model development, validation, and exploitation. *Mol Inform*. 2010;29(6–7):476–88. https://doi.org/10.1002/minf.201000061
7. Dearden J. The history and development of quantitative structure–activity relationships (QSARs). *Int J Quant Struct Relat*. 2016;1(1):1–44. https://doi.org/10.4018/IJQSPR.2016010101
8. Johnson SR. The trouble with QSAR (or how I learned to stop worrying and embrace fallacy). *J Chem Inf Model.* 2008;48(1):25–6. https://doi.org/10.1021/ci700332k
9. Verma J, Khedkar VM, Coutinho EC. 3D-QSAR in drug design—A review. *Curr Top Med Chem*. 2010;10(1):95–115. https://doi.org/10.2174/156802610790232260
10. Svetnik V, Liaw A, Tong C, Culberson JC, Sheridan RP, Feuston BP. Random forest: a classification and regression tool for compound classification and QSAR modeling. *J Chem Inf Comput Sci.* 2003;43(6):1947–58.
11. LeCun Y, Bengio Y, Hinton G. Deep learning. *Nature*. 2015;521(7553):436–44. https://doi.org/10.1038/nature14539
12. Wu Z, Pan S, Chen F, Long G, Zhang C, Philip SY. A comprehensive survey on graph neural networks. *IEEE Trans Neural Netw Learn Syst.* 2021;32(1):4–24. https://doi.org/ 10.1109/TNNLS.2020.2978386
13. Zhang L, Tan J, Han D, Zhu H. From machine learning to deep learning: progress in machine intelligence for rational drug discovery. *Drug Discov Today*. 2017;22(11):1680–5. https://doi.org/10.1016/j.drudis.2017.08.010
14. Jiménez-Luna J, Grisoni F, Schneider G. Drug discovery with explainable artificial intelligence. *Nat Mach Intell*. 2020;2(10):573–84. https://doi.org/10.1038/s42256-020-00236-4
15. Vamathevan J, Clark D, Czodrowski P, Dunham I, Ferran E, Lee G, et al. Applications of machine learning in drug discovery and development. *Nat Rev Drug Discov.* 2019;18(6):463–77. https://doi.org/10.1038/s41573-019-0024-5
16. McCulloch WS, Pitts W. A logical calculus of the ideas immanent in nervous activity. *Bull Math Biophys*. 1943;5(4):115-33. https://doi.org/10.1007/BF02478259
17. Bishop CM. Neural networks for pattern recognition. Oxford: Oxford University Press; 1995. Available from: https://people.sabanciuniv.edu/berrin/cs512/lectures/Book-Bishop-Neural%20Networks%20for%20Pattern%20Recognition.pdf
18. Hornik K, Stinchcombe M, White H. Multilayer feedforward networks are universal approximators. *Neural Netw*. 1989;2(5):359–66. https://doi.org/10.1016/0893-6080(89)90020-8
19. Nair V, Hinton GE. Rectified linear units improve restricted Boltzmann machines. In: Proceedings of the 27th International Conference on Machine Learning (ICML). 2010. p. 807–14. Available from: https://icml.cc/Conferences/2010/papers/432.pdf
20. Goodfellow I, Bengio Y, Courville A. Deep learning. Cambridge (MA): MIT Press; 2016.
21. Rumelhart DE, Hinton GE, Williams RJ. Learning representations by back-propagating errors. *Nature*. 1986;323(6088):533–6. https://doi.org/10.1038/323533a0
22. LeCun Y, Bengio Y, Hinton G. Deep learning. *Nature.* 2015;521(7553):436–44. https://doi.org/10.1038/nature14539
23. Xu Y, Ma J, Liaw A, Sheridan RP, Svetnik V. Demystifying multitask deep neural networks for quantitative structure–activity relationships. *J Chem Inf Model*. 2017;57(10):2490–504. https://doi.org/10.1021/acs.jcim.7b00087
24. Wu Z, Pan S, Chen F, Long G, Zhang C, Philip SY. A comprehensive survey on graph neural networks. *IEEE Trans Neural Netw Learn Syst*. 2021;32(1):4–24. https://doi.org/10.1109/TNNLS.2020.2978386
25. Ramsundar B, Kearnes S, Riley P, Webster D, Konerding D, Pande V. Massively multitask networks for drug discovery. *arXiv preprint arXiv*:1502.02072. 2015. https://doi.org/10.48550/arXiv.1502.02072
26. Segler MHS, Kogej T, Tyrchan C, Waller MP. Generating focused molecule libraries for drug discovery with recurrent

neural networks. *ACS Cent Sci.* 2018;4(1):120–31. https://doi.org/10.1021/acscentsci.7b00512

27. Gilmer J, Schoenholz SS, Riley PF, Vinyals O, Dahl GE. Neural message passing for quantum chemistry. In: Proceedings of the 34th International Conference on Machine Learning (ICML). 2017. p. 1263–72. https://doi.org/10.48550/arXiv.1704.01212
28. Hinton GE, Salakhutdinov RR. Reducing the dimensionality of data with neural networks. *Science*. 2006;313(5786):504–7. https://doi.org/10.1126/science.1127647
29. Gómez-Bombarelli R, Wei JN, Duvenaud D, Hernández-Lobato JM, Sánchez-Lengeling B, Sheberla D, et al. Automatic chemical design using a data-driven continuous representation of molecules. *ACS Cent Sci*. 2018;4(2):268–76. https://doi.org/ 10.1021/acscentsci.7b00572
30. Guimaraes GL, Sanchez-Lengeling B, Outeiral C, Farias PL, Aspuru-Guzik A. Objective-reinforced generative adversarial networks (ORGAN) for sequence generation models. arXiv [preprint]. 2017 May 30. Available from: https://arxiv.org/abs/1705.10843
31. Kadurin A, Nikolenko S, Khrabrov K, Aliper A, Zhavoronkov A. druGAN: an advanced generative adversarial autoencoder model for de novo generation of new molecules with desired molecular properties in silico. *Mol Pharm*. 2017;14(9):3098–104. https://doi.org/10.1021/acs.molpharmaceut.7b00346
32. Todeschini R, Consonni V. Molecular descriptors for chemoinformatics. Weinheim: Wiley-VCH; 2009. https://doi.org/10.1002/9783527628766
33. Balaban AT. Highly discriminating distance-based topological index. *Chem Phys Lett.* 1982;89(5):399–404. https://doi.org/10.1016/0009-2614(82)80009-2
34. Rogers D, Hahn M. Extended-connectivity fingerprints. *J Chem Inf Model*. 2010;50(5):742–54. https://doi.org/10.1021/ci100050t
35. Weininger D. SMILES, a chemical language and information system. *J Chem Inf Comput Sci.* 1988;28(1):31–6. https://doi.org/10.1021/ci00057a005
36. Xu Y, Dai Z, Chen F, Gao S, Pei J, Lai L. Deep learning for drug-induced liver injury. *J Chem Inf Model*. 2015;55(10):2085–93. https://doi.org/10.1021/acs.jcim.5b00238
37. Willett P. Similarity-based virtual screening using 2D fingerprints. *Drug Discov Today*. 2006;11(23–24):1046–53. https://doi.org/10.1016/j.drudis.2006.10.005
38. Rogers D, Hahn M. Extended-connectivity fingerprints. *J Chem Inf Model*. 2010;50(5):742–54. https://doi.org/10.1021/ci100050t
39. Weininger D. SMILES, a chemical language and information system. 1. Introduction to methodology and encoding rules. *J Chem Inf Comput Sci*. 1988;28(1):31–6. https://doi.org/10.1021/ci00057a005
40. Bjerrum EJ. SMILES enumeration as data augmentation for neural network modeling of molecules. arXiv preprint arXiv:1703.07076. 2017. https://doi.org/10.48550/arXiv.1703.07076
41. Veličković P, Cucurull G, Casanova A, Romero A, Lio P, Bengio Y. Graph attention networks. In: International Conference on Learning Representations (ICLR 2018); 2018.
42. Duvenaud DK, Maclaurin D, Iparraguirre J, Bombarell R, Hirzel T, Aspuru-Guzik A, Adams RP. Convolutional networks on graphs for learning molecular fingerprints. In: NeurIPS; 2015. p. 2224–32.
43. Kearnes S, McCloskey K, Berndl M, Pande V, Riley P. Molecular graph convolutions: moving beyond fingerprints. *J Comput Aided Mol Des*. 2016;30(8):595–608. https://doi.org/10.1007/s10822-016-9938-8
44. Schütt KT, Kindermans PJ, Sauceda HE, Chmiela S, Tkatchenko A, Müller KR. Quantum-chemical insights from deep tensor neural networks. *Nat Commun.* 2017;8:13890. https://doi.org/10.1038/ncomms13890
45. Thomas N, Smidt T, Kearnes S, Yang L, Li L, Kohlhoff K, et al. Tensor field networks: rotation- and translation-equivariant neural networks for 3D point clouds. arXiv [preprint]. 2018 Feb 22. Report No.: arXiv:1802.08219. https://doi.org/10.48550/arXiv.1802.08219
46. Todeschini R, Consonni V. Molecular descriptors for chemoinformatics. Weinheim: Wiley-VCH; 2009. https://doi.org/10.1002/9783527628766
47. Wu Z, Ramsundar B, Feinberg EN, Gomes J, Geniesse C, Pappu AS, et al. MoleculeNet: A benchmark for molecular machine learning. *Chem Sci*. 2018;9(2):513–30.
48. Sanchez-Lengeling B, Aspuru-Guzik A. Inverse molecular design using machine learning: generative models for matter engineering. *Science*. 2018;361(6400):360–5. https://doi.org/10.1126/science.aat2663
49. Xu Y, Ma J, Liaw A, Sheridan RP, Svetnik V. Demystifying multitask deep neural networks for Quantitative Structure-Activity Relationships. *J Chem Inf Model*. 2017;57(10):2490–504.
50. Altae-Tran H, Ramsundar B, Pappu AS, Pande V. Low data drug discovery with one-shot learning. *ACS Cent Sci*. 2017;3(4):283–93. https://doi.org/10.1021/acscentsci.6b00367
51. Caruana R. Multitask learning. *Mach Learn*. 1997;28(1):41–75. https://doi.org/10.1023/A:1007379606734
52. Ramsundar B, Eastman P, Walters P, Pande V. Deep learning for the life sciences. Sebastopol (CA): O'Reilly Media; 2019. Available from: https://www.oreilly.com/library/view/deep-learning-for/9781492039822/
53. Yang K, Swanson K, Jin W, Coley C, Eiden P, Gao H, et al. Analyzing learned molecular representations for property prediction. *J Chem Inf Model*. 2019;59(8):3370–88. https://doi.org/10.1021/acs.jcim.9b00237
54. Chemprop. Chemprop: GitHub repository. 2025. Available from: https://github.com/chemprop/chemprop
55. Jaworska J, Nikolova-Jeliazkova N, Aldenberg T. QSAR applicability domain estimation by projection of the training set descriptor space: A review. *Altern Lab Anim*. 2005;33(5):445–59. https://doi.org/10.1177/026119290503300508
56. Sheridan RP. Time-split cross-validation as a method for estimating the goodness of prospective prediction. *J Chem Inf Model*. 2013;53(4):783–90. https://doi.org/10.1021/ci400084k
57. Srivastava N, Hinton G, Krizhevsky A, Sutskever I, Salakhutdinov R. Dropout: A simple way to prevent neural networks from overfitting. *J Mach Learn Res.* 2014;15:1929-58.
58. Bergstra J, Bengio Y. Random search for hyper-parameter optimization. *J Mach Learn Res*. 2012;13:281–305.
59. Feurer M, Hutter F. Hyperparameter optimization. In: Hutter F, Kotthoff L, Vanschoren J, editors. *Automated Machine Learning*. Cham: Springer; 2019. p. 3–33.
60. Arlot S, Celisse A. A survey of cross-validation procedures for model selection. *Stat Surv*. 2010;4:40–79. https://doi.org/10.1214/09-SS054
61. Alexander DLJ, Tropsha A, Winkler DA. Beware of R2: Simple, unambiguous assessment of the prediction accuracy of QSAR and QSPR models. *J Chem Inf Model*. 2015;55(7):1316–22. https://doi.org/10.1021/acs.jcim.5b00206
62. Powers DMW. Evaluation: From precision, recall and F-measure to ROC, informedness, markedness and correlation. *J Mach Learn Technol*. 2011;2(1):37–63. https://doi.org/10.48550/arXiv.2010.16061
63. Chicco D, Jurman G. The advantages of the Matthews correlation coefficient (MCC) over F1 score and accuracy in binary classification evaluation. *BMC Genomics*. 2020;21(1):6. https://doi.org/10.1186/s12864-019-6413-7
64. Stokes JM, Yang K, Swanson K, Jin W, Cubillos-Ruiz A, Donghia NM, et al. A deep learning approach to antibiotic discovery. *Cell*. 2020;180(4):688–702.e13. https://doi.org/10.1016/j.cell.2020.01.021
65. Schneider G. Automating drug discovery. *Nat Rev Drug Discov*. 2018;17(2):97–113. https://doi.org/10.1038/nrd.2017.232
66. Polishchuk P. Interpretation of quantitative structure–activity relationship models: Past, present, and future. *J Chem Inf Model*. 2017;57(11):2618–39. https://doi.org/10.1021/acs.jcim.7b00274
67. Helma C, Cramer T, Kramer S, De Raedt L. Data mining and machine learning techniques for the identification of mutagenicity inducing substructures and structure activity relationships of noncongeneric compounds. *J Chem Inf Comput Sci.* 2004;44(4):1402–11. https://doi.org/10.1021/ci034254q

68. Ramsundar B, Liu B, Wu Z, Verras A, Tudor M, Sheridan RP, et al. Is multitask deep learning practical for pharma? *J Chem Inf Model*. 2017;57(8):2068–76. https://doi.org/10.1021/acs.jcim.7b00146

69. Chen B, Ma L, Paik H, Sirota M, Wei W, Chua MS, et al. Reversal of cancer gene expression correlates with drug efficacy and reveals therapeutic targets. *Nat Commun*. 2017;8:16022. https://doi.org/10.1038/ncomms16022

70. Sharifi-Noghabi H, Zolotareva O, Collins CC, Ester M. MOLI: Multi-omics late integration with deep neural networks for drug response prediction. *Bioinformatics*. 2019;35(14):i501–9. https://doi.org/10.1093/bioinformatics/btz318

71. Iorio F, Knijnenburg TA, Vis DJ, Bignell GR, Menden MP, Schubert M, et al. A landscape of pharmacogenomic interactions in cancer. *Cell*. 2016;166(3):740–54. https://doi.org/10.1016/j.cell.2016.06.017

72. Hodos R, Xu J, Kim H, Resnick E, Skorecki K, Salomonis N, et al. Integrative network-based approaches for precision medicine. *Curr Opin Syst Biol*. 2018;9:17–25.

73. OECD. Principles for the validation, for regulatory purposes, of (quantitative) structure–activity relationship models. Paris: OECD Publishing; 2004. Available from: https://www.oecd.org/content/dam/oecd/en/topics/policy-sub-issues/assessment-of-chemicals/oecd-principles-for-the-validation-for-regulatory-purposes-of-quantitative-structure-activity-relationship-models.pdf

74. Fourches D, Muratov E, Tropsha A. Trust, but verify: On the importance of chemical structure curation in cheminformatics and QSAR modeling research. *J Chem Inf Model*. 2010;50(7):1189–204. https://doi.org/10.1021/ci100176x

75. Gaulton A, Hersey A, Nowotka M, Bento AP, Chambers J, Mendez D, et al. The ChEMBL database in 2017. *Nucleic Acids Res*. 2017;45(D1):D945–54. https://doi.org/10.1093/nar/gkw1074

76. Zhavoronkov A, Ivanenkov YA, Aliper A, Veselov MS, Aladinskiy VA, Aladinskaya AV, et al. Deep learning enables rapid identification of potent DDR1 kinase inhibitors. *Nat Biotechnol*. 2019;37(9):1038–40. https://doi.org/10.1038/s41587-019-0224-x

77. Biamonte J, Wittek P, Pancotti N, Rebentrost P, Wiebe N, Lloyd S. Quantum machine learning. Nature. 2017;549(7671):195–202.

Cite this article: Shaikh S, Alte A, Pande V, Wagh V, Suryawanshi J. Artificial Intelligence in Modern Deep-Learning QSAR. *J Educ Technol Health Sci.* 2025;12(3):77–88.