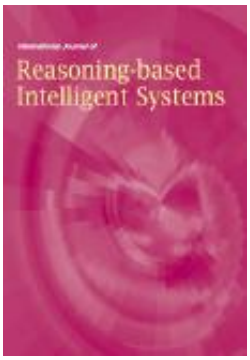


International Journal of Reasoning-based Intelligent Systems

ISSN online: 1755-0564 - ISSN print: 1755-0556
https://www.inderscience.com/ijris

Deep learning-based virtual screening system for drug molecules

Chuyue Zhang

DOI: 10.1504/IJRIS.2026.10076055

Article History:

Received:	16 November 2025
Last revised:	19 December 2025
Accepted:	20 December 2025
Published online:	17 February 2026

Deep learning-based virtual screening system for drug molecules

Chuyue Zhang

College of Pharmacy,
China Pharmaceutical University,
Nanjing, 211198, China
Email: 2020231291@stu.cpu.edu.cn

Abstract: In the field of drug discovery, traditional virtual screening methods face challenges of being time-consuming, costly, and limited in accuracy. To address this, this study developed a deep learning-based virtual screening system for drug molecules. By automatically learning key molecular features through graph neural networks, the system overcomes the limitations of traditional methods that rely on manual feature extraction, thereby capturing more complex structural information. Testing on the public benchmark Directory of Useful Decoys: Enhanced (DUD-E) demonstrates that this system achieves an area under the curve (AUC) of 0.889 while significantly reducing screening time – approximately 80% faster than conventional methods. This research provides an efficient solution for rapidly and accurately identifying potential drug candidates from vast compound libraries, paving the way for accelerated drug development.

Keywords: virtual screening; deep learning; drug discovery; graph neural networks; molecular characterisation.

Reference to this paper should be made as follows: Zhang, C. (2026) 'Deep learning-based virtual screening system for drug molecules', *Int. J. Reasoning-based Intelligent Systems*, Vol. 18, No. 8, pp.44–55.

Biographical notes: Chuyue Zhang entered China Pharmaceutical University in 2023 and she is currently a student majoring in Pharmaceutical Preparations. Her research interests include pharmaceutical preparations, data mining and statistical analysis.

1 Introduction

Drug discovery is a serious and highly demanding endeavour that is crucial for combating human diseases and extending lifespan. The conventional new drug development processes are time consuming, in most cases, they take over ten years and involve very huge investments amounting to tens of billions of dollars. The major causes of this cost and time are the screening and optimisation of candidate compounds (Estrela and Hemanth, 2015). Confronted with a large chemical space that has millions to billions of compounds, the classical methods of experimental screening such as high-throughput screening are limited by physical throughput even in highly automated form, and the cost of these technologies is prohibitive of extensive sampling of the chemical space. It is against this backdrop that computer-aided drug design, especially virtual screening has become an inevitable part of an up to date drug discovery process (Hemanth and Estrela, 2013). Virtual screening is selective based on the large molecular databases, utilising computational models to predict those molecules that are most likely to bind to particular disease targets (a protein). This will focus the limited resources in experiments within the best candidate molecules, thereby significantly increasing the efficiency of research and development.

Despite this potential, the practical effectiveness of virtual screening remains constrained by the inherent limitations of its classical computational approaches. Virtual screening methods can be broadly categorised into two types: receptor-based and ligand-based approaches. Receptor-based methods, such as molecular docking, rely on the three-dimensional structural information of target proteins (Wang et al., 2024). They predict binding affinity by simulating the geometric and chemical complementarity between small molecules and target binding sites. However, this approach heavily relies on high-resolution protein structures, which may remain unresolved or be of poor quality for many important drug targets. Additionally, issues such as conformation search and scoring function accuracy during computation limit its predictive precision (Matsumura et al., 2025). On the other hand, ligand-based approaches, such as pharmacophore models and quantitative structure-activity relationship (QSAR) models, screen new compounds by analysing common structural features of known active molecules when receptor structural information is lacking. While computationally efficient, the performance of these methods heavily depends on the quantity and quality of known active compounds (Ishida et al., 2023). They often require expert knowledge for manual feature extraction, struggle to capture complex and subtle structure-activity relationships between molecules, and exhibit limited generalisation capabilities.

Consequently, a fundamental challenge shared by both receptor-based and ligand-based paradigms is their dependence on explicit, pre-defined rules or expert-crafted features, which may not fully capture the complex nature of biomolecular interactions. To overcome the limitations of traditional methods, machine learning techniques have been introduced into the field of virtual screening. The initial research used the classical machine learning algorithms like support-vector machines, random-forests, and naive Bayes to illustrate better performance compared to the traditional methods on a particular dataset. Nevertheless, these models still rely on the manually computed molecular descriptors (e.g., fingerprints, physicochemical properties). Not only does the construction of these descriptors demand specialised domain knowledge, but it can equally contribute unintentionally to designer bias missing complex structural details that may be important to the biological activity during the feature engineering phase (Esakia et al., 2025). Notably, the inherent representation of a molecule – a graph where atoms are nodes and bonds are edges – provides a natural and rich substrate for automated feature learning, directly addressing this bottleneck. The topological graph structure of molecules can be considered simply as the atomic structure, that is, atoms can be seen as nodes and chemical bonds can be seen as edges. The mechanism of empowering machines to automatically and effectively develop the very depth of such graph structures is the success variable in increasing the virtual screening effectiveness (Zhu et al., 2024).

Over the last few years, deep learning has evolved at a very rapid pace and offered a new paradigm through which these problems could be resolved. Specifically, the appearance of graph neural networks, which are powerful neural nets that are specifically trained on graph-structured data, allows transitioning to directly learning the raw graph representations of molecules, without having to go through the tedious and possibly biased manual engineering of features. As an example, atom graphs convolutional networks and variants may be trained to learn distributed representations of the entire molecule, through the representation of each atom by continually summing the representation of connected atoms, thus learning distributed representations of all atoms (Li et al., 2025). Breaking through methods have shown the potential of graph neural network (GNN) in the prediction of molecular properties and activities, which outperform machine learning methods on multiple public benchmark systems based on traditional descriptors (Hosseini et al., 2024). Moreover, convolutional neural networks and recurrent neural networks have also been examined at processing the linear representations of molecules, e.g., simplified molecular input line entry system (SMILES) strings, with success. These developments mark the transition of virtual screening to the period of representation learning and not feature engineers.

Despite its promising prospects, the comprehensive and effective application of deep learning to large-scale virtual screening of drug molecules still faces several significant challenges (Kim et al., 2024). First, most existing deep learning models are designed for a single type of data (e.g., graph structures or SMILES strings), failing to fully leverage the complementary information contained in multi-source heterogeneous data (e.g., combining molecular graph information with known drug-target interaction networks). This limitation may constrain the models' predictive power and generalisation ability (Ferreo et al., 2022). Second, drug chemistry data frequently suffers from significant data sparsity – the number of molecules with known activity is vastly smaller than the compound library being screened – posing an overfitting risk for deep model training. Furthermore, many deep learning models operate as 'black boxes', lacking intuitive chemical or biological explanations for their predictive decisions (Lam et al., 2025). This makes it difficult for medicinal chemists to trust model outputs and use them for rational molecular design optimisation. Finally, building an end-to-end automated system that seamlessly integrates data preprocessing, model training, large-scale inference, and result visualisation is crucial for transforming advanced deep learning technologies into user-friendly tools for drug discovery scientists. Systematic work in this area remains to be explored in depth. Therefore, developing a novel deep learning framework that effectively addresses these challenges holds significant theoretical value and practical significance for advancing virtual screening technology and accelerating the drug discovery process.

2 Research review

2.1 Traditional virtual screening methods and their limitations

The early development and core paradigm of virtual screening heavily relied on principles from computational chemistry and structural biology (Ann and William, 2025). Receptor-based virtual screening, also known as molecular docking, predicts binding patterns and affinities by computationally assessing the complementarity between small-molecule ligands and biomolecular targets in three-dimensional space. Its physical basis is typically measured by the binding free energy ΔG_{bind} and estimated through empirical scoring functions. A general form can be expressed as:

$$Score = E_{\text{interaction}} + E_{\text{internal}} + \Delta G_{\text{solvation}} + ...,$$

where $E_{\text{interaction}}$ represents interaction energies between the ligand and receptor, such as van der Waals and electrostatic forces, E_{internal} denotes the internal strain energy of the ligand molecule, and $\Delta G_{\text{sovation}}$ signifies the free energy change due to solvation effects. Despite decades of refinement, these scoring functions still face significant challenges in handling flexible ligands, water-mediated interactions, and entropy change estimation, leading to substantial discrepancies between their predictions and experimental measurements (Venkatesan et al., 2025). On the other hand, ligand-based screening methods do not rely on the target's three-dimensional structure. Their core

assumption is that structurally similar molecules may exhibit similar biological activity. These methods depend on predefined molecular descriptors or fingerprints [such as molecular access system (MACCS) keys or extended-connectivity fingerprint (ECFP) fingerprints] to calculate molecular similarity. Common metrics include the Tanimoto coefficient:

$$T(A,B)=\frac{|F_A \cap F_B|}{|F_A \cup F_B|},$$

where F_A and F_B represent the feature fingerprint vectors of molecules A and B, respectively. However, this 'principle of similarity' has well-known exceptions, where minor structural alterations can lead to drastic loss or gain of activity. More critically, both receptor- and ligand-based traditional approaches heavily rely on expert prior knowledge. Subjective biases may be introduced throughout the process – from protein preparation and active site definition to molecular descriptor selection – limiting their applicability for exploring unknown targets or entirely novel chemical spaces. These limitations have motivated the integration of data-driven machine learning techniques to build more adaptive and predictive virtual screening models.

2.2 The evolution of machine learning in virtual screening

To reduce reliance on manual expertise, machine learning models have been widely adopted to establish quantitative or qualitative mapping relationships between molecular features and biological activity. Support vector machines construct a hyperplane to separate active and inactive molecules, with a decision function defined as:

$$f(x)=\text{sign}\left(\sum_{i=1}^{n}\alpha_i y_i K(x_i,x)+b\right),$$

where α_i is the Lagrange multiplier, y_i is the sample label, $K(x_i, x)$ is the kernel function, which maps input features x to a high-dimensional space to achieve linear separability. Random forests perform predictions by integrating multiple decision trees, with the final result determined by the mode (for classification) or mean (for regression) of all trees' outputs:

$$\hat{y}=\frac{1}{B}\sum_{b=1}^{B}T_b(x),$$

where B is the number of trees and $T_b(x)$ is the prediction of the b^{th} tree. While these methods enhance virtual screening accuracy to some extent, they fundamentally remain within the paradigm of 'feature engineering'. The process of model performance is closely linked to molecular descriptors, and the construction of the descriptors is an information bottleneck. The thing is that complex chemical intuitions and stereochemical information, not easy to be formalised mathematically but important to the activity, are probably lost in the process of feature extraction. This establishes a strong ceiling effect on models and causes unstable generalisation between various datasets. The pursuit of overcoming this performance ceiling and automating the feature discovery process naturally led to the exploration of deep representation learning.

2.3 Deep learning and molecular representation learning

The revolutionary breakthrough of deep learning lies in its ability to automatically learn discriminative hierarchical feature representations from raw data. In virtual drug screening, this means models can learn directly from raw molecular representations – such as graph structures or SMILES strings – bypassing the bottleneck of manually designing descriptors. For data treating molecules as graph structures, graph neural networks emerge as a natural and powerful choice (Gilmer et al., 2017). Graph convolutional networks represent one of the most prominent architectures, where a single layer's message-passing mechanism can be expressed as:

$$H^{(l+1)}=\sigma\left(\tilde{D}^{-\frac{1}{2}}\tilde{A}\tilde{D}^{-\frac{1}{2}}H^{(l)}W^{(l)}\right),$$

where $\tilde{A}=A+I$ is the adjacency matrix augmented with self-loops, $\tilde{D}$ is the degree matrix of $\tilde{A}$, $H^{(l)}$ is the node feature matrix at layer l, $W^{(l)}$ is the trainable weight matrix for that layer, and σ is a nonlinear activation function. Through multi-layer stacking, the final representation of each atom (node) can capture both its local chemical environment and the global structural information of the entire molecular graph. Based on this, the representation of the entire molecule (graph-level representation) h_G can be obtained through a global pooling operation, such as using the mean of all node features:

$$h_G=\frac{1}{N}\sum_{i=1}^{N}h_i^{(L)},$$

where N is the number of nodes, L denotes the final layer, and $h_i^{(L)}$ represents the final feature of atom i. Beyond GCNs, graph attention networks introduce an attention mechanism that enables weighted aggregation of neighbour information, where weights are determined by the node's own features:

$$\alpha_{ij}=\frac{\exp\left(\text{LeakyReLU}\left(a^T\left[Wh_i\|Wh_j\right]\right)\right)}{\sum_{k\in\mathcal{N}_i}\exp\left(\text{LeakyReLU}\left(a^T\left[Wh_i\|Wh_k\right]\right)\right)}.$$

Here, a is the attention vector, W is the weight matrix, and || denotes the concatenation operation. This mechanism enables the model to focus on chemical substructures that are more important for the task. On the other hand, for sequential data such as SMILES strings, recurrent neural networks and their variants like long short-term memory networks can handle sequential dependencies, where the

update of cell states involves the input gate i_t, the forget gate f_t, the output gate o_t and the candidate cell state $\tilde{C}_t$:

$$C_t = f_t \odot C_{t-1} + i_t \odot \tilde{C}_t,\ h_t = o_t \odot \tanh(C_t).$$

These advanced deep learning architectures enable genuine molecular representation learning, laying the foundation for constructing smarter virtual screening systems. Building such 'smarter' systems further benefits from integrating these learned molecular representations with complementary biological data, a direction explored in multimodal fusion approaches.

2.4 Multimodal data fusion and model optimisation for virtual screening

Although single-molecule representation learning has achieved significant success, the complexity of drug discovery demands models capable of integrating multi-source heterogeneous information (Albarillo and Fernandez, 2023). For instance, combining a molecule's graph structural information with a target protein's sequence or structural data can simulate biophysical interactions closer to reality. Multimodal deep learning frameworks achieve this by designing cross-attention mechanisms or joint embedding spaces. For a molecule-target pair, their joint representation can be modelled via bilinear interaction:

$$s = h_{\text{molecule}}^T M h_{\text{target}},$$

where h_{molecule} and h_{targer} are the embedding vectors for the molecule and target, respectively, and M is a learnable interaction matrix. Furthermore, the inherent sparsity and class imbalance in medicinal chemistry data pose significant challenges for deep learning models. To mitigate overfitting risks and enhance model generalisation, advanced regularisation techniques like dropout are widely adopted. During training, dropout randomly 'drops' neurons with a predefined probability p:

$$h' = h \odot \text{mask},\ \text{mask} \sim \text{Bernuoulli}(1-p).$$

Simultaneously, transfer learning strategies demonstrate significant potential. This involves training a base model on a large, general molecular property dataset [e.g., chemistry of bioactive molecules (ChEMBL)] to acquire a universal understanding of chemical space, then fine-tuning this model using small-scale activity data for a specific target:

$$\theta^* = \arg\min_\theta \mathcal{L}\text{fine-tuning}\left(f\theta\left(X_{target}\right), Y_{target}\right),$$

where model parameters θ are initialised using parameters learned on the pretraining task. This 'pretrain-fine-tune' paradigm significantly alleviates data scarcity issues.

2.5 Research progress on model interpretability

One of the main obstacles to apply deep learning models in high-risk drug discovery is the skepticism that the models provide a black box (Johannes, 2022). Thus, it is essential to improve the model interpretability. Posterior explanation-based techniques, including gradient-based attribution techniques, determine the importance of features by computing the gradient of the model outputs with respect to the input features:

$$I_i = \left|\frac{\partial y}{\partial x_i}\right|.$$

For graph models, this yields the contribution of each atom and chemical bond pair to the prediction outcome. Furthermore, attention mechanisms inherently provide built-in interpretability. The attention weights α_{ij} in graph attention network (GAT), which, as stated above, intuitively provide information on the key interactions between atomic pairs or chemical bond interactions model takes into account when passing messages. These methods will convert the model predictions into a form intelligible to medicinal chemists like the ability to highlight the molecular substructures that are important in predicting activity and they will raise the confidence of a researcher in the model and maybe even enable rational design information about the molecular optimisation to follow. The pursuit of explainable AI models remains an active and critical area of research for their broader adoption in high-stakes domains like drug discovery (Yang et al., 2019).

3 Research methods

In this section, we will describe in detail the architecture and the core algorithms of an end-to-end deep learning-based virtual screening system for drug molecules. The goal of this architecture is to provide a predictive workflow from the raw input of molecular data to the output of the activity probability. The core contribution of our architecture is the multimodal molecular characterisation combined with a novel graph neural network architecture.

3.1 Overall system framework

The virtual screening system proposed in this study is an integrated computational platform, with its core workflow illustrated in Figure 1. The system accepts input in the form of standardised molecular structure files (e.g., structure data file files) or SMILES strings. First, the data preprocessing module standardises molecules, removes salts, neutralises charges, and uniformly converts them into internal graph representations. Then, processed data of molecular graphs are employed into a multi-channel molecular graph neural network to extract features. Simultaneously, this network is working with molecular topology, atomic physicochemical properties and syntactic aspects obtained with SMILES sequences. The high-dimensional features that are extracted are then processed by a multimodal feature fusion module to incorporate and refine the features. Lastly, the merged features are used to feed a multi-layer perceptron (MLP) classifier that makes the determination of a predicted probability of a molecule binding activity in response to a particular target. The whole system is developed and

executed on a high-performance computing cluster based on a number of graphics processing unit (GPU) and thus efficient in processing big databases of proteins. The following subsections detail the key components of this framework: molecular representation (Section 3.2), the multi-channel GNN architecture (Section 3.3), and the multimodal fusion mechanism (Section 3.4).

3.2 Molecular representation and initialisation

An effective representation of molecules is fundamental to the success of a model. This system employs a graph structure $G = (V, E)$ – where atoms form nodes and chemical bonds form edges – as the primary representation for molecules. Here, V denotes the set of nodes (atoms), and E denotes the set of edges (chemical bonds).

Each atomic node $v_i \in V$ is initialised with a feature vector x_i encoding the atom's fundamental properties:

$$x_i = \left[\mathrm{Emb}(a_i); \deg(v_i); \mathrm{val}(v_i); \mathrm{hyb}(v_i); \mathrm{aroma}(v_i)\right] \quad (1)$$

where $\mathrm{Emb}(a_i)$ is the dense vector learned by the embedding layer for atom type a_i (e.g., C, N, O); $\deg(v_i)$ is the degree (connection count) of atom v_i; $\mathrm{val}(v_i)$ is the valence electron count of the atom; $\mathrm{hyb}(v_i)$ is the hybridisation state of the atom; $\mathrm{aroma}(v_i)$ is a binary value indicating whether the atom belongs to an aromatic ring. These features collectively form the initial state h_i^0 of the atom.

For each chemical bond edge $e_{ij} \in E$, initialise with a single eigenvector e_{ij} as follows:

$$e_{ij} = \left[\mathrm{Emb}(b_{ij}); \mathrm{conj}(e_{ij}); \mathrm{ring}(e_{ij})\right] \quad (2)$$

where $\mathrm{Emb}(b_{ij})$ denotes the embedding vector for bond type b_{ij} (e.g., single bond, double bond, triple bond); $\mathrm{conj}(e_{ij})$ indicates whether the bond is in a conjugated system; $\mathrm{ring}(e_{ij})$ indicates whether the bond is within a ring structure.

3.3 Multi-channel graph neural network architecture

To comprehensively capture the complex information of molecules, we designed a multi-channel graph neural network. At its core lies an improved graph attention network, and we introduced residual connections to mitigate the vanishing gradient problem in deep networks.

3.3.1 Messaging and attention mechanisms

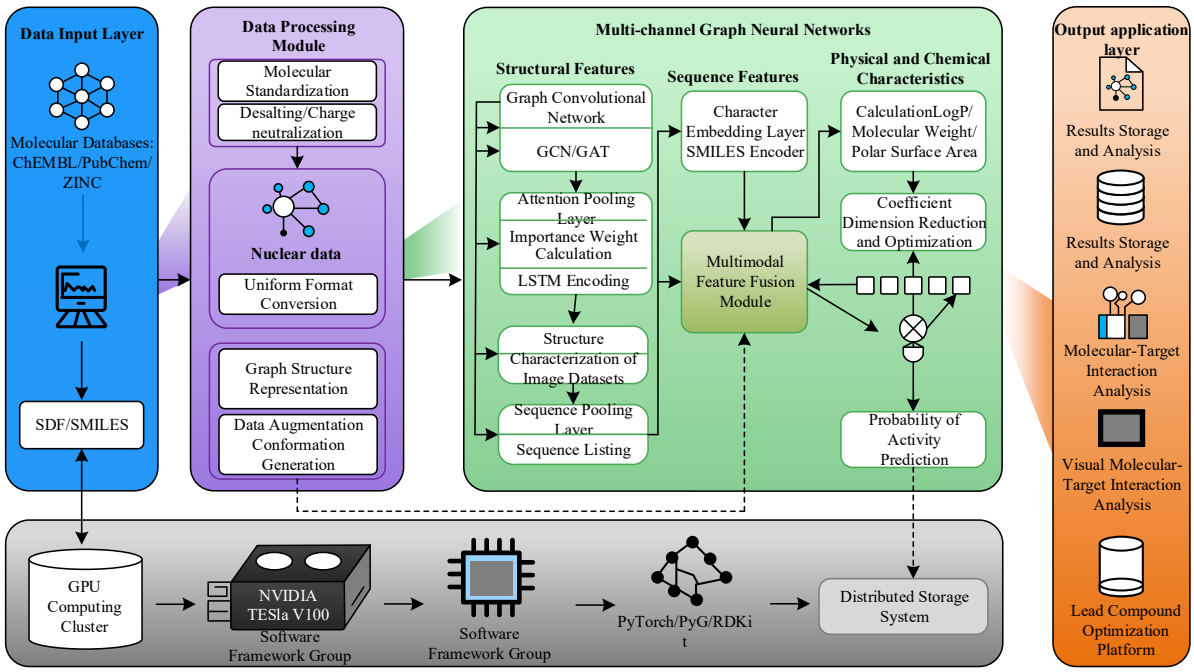
At layer l, for atom i, messages transmitted by its neighbouring atom $j \in \mathcal{N}(i)$ are weighted and aggregated through an attention mechanism conditioned on edge features. First, the attention coefficient α_{ij} from atom j to atom i is computed:

$$e_{ij}^l = \mathrm{LeakyReLU}\left(W_a^l \left[h_i^l \middle| h_j^l \middle| e_{ij}\right]\right) \quad (3)$$

$$\alpha_{ij}^l = \frac{\exp\left(e_{ij}^l\right)}{\sum_{k \in \mathcal{N}(i)} \exp\left(e_{ik}^l\right)} \quad (4)$$

where W_a^l denotes the trainable attention weight matrix for layer l, $[\cdot|\cdot]$represents vector concatenation, h_i^l and h_j^l are the features of atoms i and j in layer l, respectively, and e_{ij} is the edge feature. α_{ij}^l represents the normalised attention weight, which determines the importance of information from neighbouring atom j when aggregated into atom i.

Figure 1 Architecture diagram of a deep learning-based virtual screening system for drug molecules (see online version for colours)



3.3.2 Feature update

Subsequently, the feature of atom i is updated by aggregating messages from all its neighbours and combining them with its own information. We employ a multi-head attention mechanism to stabilise the learning process:

$$\tilde{h}i^{l} = \Big|_{m=1}^{M} \sigma\left(\sum_{j \in \mathcal{N}(i)} \alpha_{ij}^{l,m} \cdot W_{v}^{l,m} h_{j}^{l} \right) \tag{5}$$

$$h_{i}^{l+1} = \mathrm{LayerNorm}\left(h_{i}^{l} + \mathrm{Dropout}\left(W_{o}^{l} \tilde{h}_{i}^{l} \right) \right) \tag{6}$$

where M denotes the number of attention heads, | represents vector concatenation, σ is the exponential linear unit (ELU) activation function, $W_{v}^{l,m}$ is the value transformation matrix corresponding to the m^{th} head, and W_{o}^{l} is the output matrix used to map the concatenated features back to the original dimension. LayerNorm denotes layer normalisation, while dropout prevents overfitting. This update formula ensures each atom can simultaneously perceive its local chemical environment while preserving its own properties.

This iterative update process allows each atom's representation to be dynamically refined based on its local chemical environment, which is crucial for capturing substructural features relevant to biological activity. Architectural configuration. Our improved GAT consists of 4 graph attention layers. The number of attention heads K was set to 8 for the first two layers and 4 for the last two layers, with each head having a dimensionality of 64. The resulting node features were therefore 512-dimensional after the first two layers and 256-dimensional after the last two. The dropout rate was set to 0.2, and the ELU activation function was used throughout. The ELU function is defined as:

$$f(x) = \begin{cases} x & \text{if } x > 0 \\ \alpha\left(e^{x} - 1\right) & \text{if } x \le 0 \end{cases}$$

with α = 1.0. This was chosen over the common rectified linear unit (ReLU) function ($f(x)$ = max(0, x)) to mitigate the 'dying ReLU'. Residual connections were added from the input to the output of each graph attention layer to facilitate the training of this deeper architecture.

3.4 Multimodal feature fusion and graph-level representation

After L iterations of the GNN, we obtain the final feature h_{i}^{L} for each atom, rich in contextual information. To make predictions for the entire molecule, all atomic features must be aggregated into a global graph-level representation h_G. We employ an attention pooling mechanism that enables the model to automatically learn each atom's contribution to the global task:

$$\beta_{i} = \mathrm{sigmoid}\left(U_{\text{attn}}^{T} h_{i}^{L} + b_{\text{attn}} \right) \tag{7}$$

$$h_{G}^{(\text{struct})} = \sum_{i=1}^{N} \beta_{i} \cdot h_{i}^{L} \tag{8}$$

where U_{attn} and b_{attn} are trainable weight vectors and bias terms, respectively; β_i is the importance weight for atom i; and $h_{G}^{(\text{struct})}$ is the molecular representation obtained from the topological structure channel.

To further enhance the model's expressive power, we introduced a SMILES sequence channel. A bidirectional long short-term memory (LSTM) network processes the SMILES strings of molecules:

$$\vec{h}t, \vec{c}t = \mathrm{LSTMforward}\left(\mathrm{Emb}\left(s_{t} \right), \vec{h}t-1, \vec{c}t-1 \right) \tag{9}$$

$$\overleftarrow{h}t, \overleftarrow{c}t = \mathrm{LSTMbackward}\left(\mathrm{Emb}\left(s_{t} \right), \overleftarrow{h}t+1, \overleftarrow{c}t+1 \right) \tag{10}$$

$$h_{t} = \left[\vec{h}_{t} \middle| \overleftarrow{h}_{t} \right] \tag{11}$$

$$h_{G}^{(\text{seq})} = \mathrm{MaxPooling}\left(\left[h_{1}, h_{2}, \ldots, h_{T} \right] \right) \tag{12}$$

where s_t denotes the t^{th} character in the SMILES sequence, Emb(s_t) represents its embedding vector, h_t is the bidirectional hidden state concatenation at time step t, and $h_{G}^{(\text{seq})}$ is the sequence-channel molecular representation obtained via max pooling.

3.4.1 Sequence channel parameters

The bidirectional LSTM for processing SMILES sequences has 2 layers with a hidden state size of 256 per direction, resulting in a final sequence representation $\mathbf{h}_{\text{seq}}$ of 512 dimensions. The character embedding dimension was set to 128. The SMILES strings were tokenised using a character-level tokeniser from the RDKit library, and the maximum sequence length was capped at 256, with shorter sequences being padded.

Finally, merge the molecular representations of the two channels:

$$h_{G} = W_{\text{fuse}} \left[h_{G}^{(\text{struct})} \middle| h_{G}^{(\text{seq})} \right] + b_{\text{fuse}} \tag{13}$$

where W_{fuse} and b_{fuse} represent the weights and biases of the fusion layer, while h_G denotes the final fused molecular representation.

3.5 Activity prediction and model training

The fused molecular feature h_G is fed into a MLP for final activity classification:

$$z_{1} = \mathrm{Dropout}\left(\mathrm{ReLU}\left(W_{1} h_{G} + b_{1} \right) \right) \tag{14}$$

$$z_{2} = \mathrm{Dropout}\left(\mathrm{ReLU}\left(W_{2} z_{1} + b_{2} \right) \right) \tag{15}$$

$$\hat{y} = \mathrm{sigmoid}\left(W_{3} z_{2} + b_{3} \right) \tag{16}$$

where W_1, W_2, W_3 and b_1, b_2, b_3 are the parameters of each layer in the MLP, and $\hat{y} \in [0, 1]$ represents the probability predicted by the model that the molecule exhibits activity.

The training objective of the model is to minimise the loss between the predicted values and the true labels y. Since virtual screening data often suffers from positive-negative sample imbalance, we employ a weighted binary cross-entropy loss function:

$$\mathcal{L} = -\frac{1}{N}\sum_{i=1}^{N}\left[w \cdot y_i \log\left(\hat{y}_i\right) + \left(1-y_i\right)\log\left(1-\hat{y}_i\right)\right] \tag{17}$$

where N denotes the number of samples in a batch, while w represents the weight assigned to positive samples to balance class imbalance. Its value is typically set as the ratio of negative to positive samples in the dataset.

To prevent overfitting, we incorporated an L2 regularisation term into the loss function:

$$\mathcal{L}_{\text{total}} = \mathcal{L} + \lambda|\Theta|_2^2 \tag{18}$$

where λ is the regularisation coefficient, and Θ represents the set of all weight parameters in the model.

The model is trained using the adaptive moment estimation with weight decay (AdamW) optimiser, which decouples weight decay and typically achieves better generalisation performance. Its parameter update rules are as follows:

$$m_t = \beta_1 m_{t-1} + \left(1-\beta_1\right) g_t \tag{19}$$

$$v_t = \beta_2 v_{t-1} + \left(1-\beta_2\right) g_t^2 \tag{20}$$

$$\hat{m}_t = \frac{m_t}{1-\beta_1^t}, \quad \hat{v}t = \frac{v_t}{1-\beta_2^t} \tag{21}$$

$$\theta_t = \theta t - 1 - \eta\left(\frac{\hat{m}_t}{\sqrt{\hat{v}t}+\varepsilon} + \lambda\theta t - 1\right) \tag{22}$$

where g_t denotes the gradient at time step t, m_t and v_t represent the first- and second-order moment estimates, respectively, β_1 and β_2 are the exponential decay rates for the moment estimates, η is the learning rate, ε is a threshold value to prevent division-by-zero errors, and λ is the weight decay coefficient.

3.5.1 Training hyperparameters

The MLP classifier consists of three linear layers with dimensions [512, 256, 128] and a final output layer of dimension 1. ReLU activation ($f(x)$ = max(0, x)) and batch normalisation were applied between the linear layers. The model was trained for a maximum of 500 epochs using the AdamW optimiser with a learning rate of 10^{-3}, betas (β_1, β_2) = (0.9, 0.999), epsilon $\varepsilon = 10^{-8}$, and a weight decay $\lambda = 10^{-4}$. We employed a batch size of 128 and used a reduce learning rate on plateau scheduler which reduced the learning rate by a factor of 0.5, if the validation loss did not improve for 15 consecutive epochs. This adaptive learning rate scheduling strategy helps refine convergence by allowing the model to navigate more precisely toward a local minimum when progress stagnates. Early stopping with a patience of 30 epochs was used to prevent overfitting.

4 Empirical research

To comprehensively evaluate the performance of the multi-channel graph neural network for virtual screening system proposed in this study (referred to as MC-GNNVS), we designed and conducted a series of rigorous experiments. The following are the main questions that were also to be answered in these experiments:

1 Does MC-GNNVS show a better predictive performance than the present conventional approaches using standard benchmark datasets?

2 What are the main elements of the model (e.g., multimodal fusion, attention pooling) and how they help improve the end performance?

3 Can the model compute effectively to the large-scale needs of screening?

4.1 Experimental setup

4.1.1 Dataset and preprocessing.

Two popular public benchmark datasets that are commonly used in the realm of drug discovery were used in this work; Directory of Useful Decoys: Enhanced (DUD-E) and a subset of ChEMBL. These datasets are widely adopted benchmarks in the field of molecular machine learning (Wu et al., 2018).

- DUD-E (Mysinger et al., 2012): this data consists of 102 protein targets, with each sample having experimentally measured active (positive samples) and decoys (negative samples) molecules produced in a systematic manner. The properties of these decoys are similar in physicochemical but they have different topological structures. This construction is effective in terms of establishing the talent of a model in identifying the finer differences in structure. To ensure a rigorous assessment of model generalisation and prevent optimistic bias from scaffold leakage, we employed a scaffold-based split for all datasets. The detailed splitting protocol is described in the following 'Data splitting and reproducibility' subsection.
- ChEMBL: we have chosen five large drug targets (added proteins such as kinases, GPCRs, etc.) of the activity data of the ChEMBL 28 database and have combined the activity data (IC50 or Ki 10,000 ml was considered active). With an aim of creating negative samples, we used a random sampling strategy among the inactive molecules of ZINC Is Not Commercial (ZINC)15 database where the ratio of positive to the negative samples was 1:40 to replicate the practicing virtual screening context in the real world.

All molecules were standardised using RDKit, including desalting, generating tautomers, and neutralising charges. Finally, molecules were uniformly converted into SMILES strings and graph structures for model input.

4.1.1.1 Data splitting and reproducibility

To prevent data leakage and ensure a realistic assessment of generalisation to novel chemotypes, a strict scaffold-based splitting strategy was employed for all experiments reported in this paper. Specifically, for both DUD-E and ChEMBL datasets, we used the Bemis-Murcko scaffold decomposition as implemented in RDKit. The data was split at the scaffold level into training (70%), validation (10%), and test (20%) sets, guaranteeing that molecules sharing a common Bemis-Murcko scaffold were exclusively contained within one of the splits. This scaffold-based splitting approach is critical for avoiding artificially inflated performance metrics due to analogue bias, and is considered a best practice for rigorous virtual screening evaluation (Truchon and Bayly, 2007). Specifically, we used the Bemis-Murcko scaffold decomposition as implemented in RDKit and performed a 70/10/20 split (train/validation/test) at the scaffold level. This means that molecules sharing a common Bemis-Murcko scaffold were exclusively contained within one of the splits. The random seed for all stochastic processes (e.g., data splitting, model weight initialisation) was fixed to 42 for reproducibility. The exact version of the datasets used is DUD-E v2.0 and ChEMBL version 28.

4.1.2 Comparison algorithm

To ensure a fair comparison, we selected the following representative methods as baselines and rigorously reproduced their experimental settings as described in the original literature:

1 Machine learning methods based on traditional descriptors
   - SVM-RBF: support vector machine with ECFP4 molecular fingerprint as features and radial basis function kernel.
   - RF: random forest model constructed using ECFP4 molecular fingerprint and physicochemical descriptors (e.g., molecular weight, LogP) as inputs.

2 Deep learning-based sequence/graph models
   - DeepDTA: a deep learning method based on convolutional neural networks for processing protein sequences and molecular SMILES strings. We adapted it to predict molecular activity using only SMILES input.
   - GCN: a standard graph convolutional network that takes atomic features and adjacency matrices as input, followed by global average pooling and a classifier.
   - GAT: which aggregates neighbouring information through an attention mechanism, is one of the core components of this model.
   - Attentive FP: an advanced graph neural network that employs attention mechanisms to simultaneously process representations of both individual atoms and the entire graph.

4.1.2.1 Baseline implementation details

All baseline models were implemented and rigorously tuned to ensure a fair comparison. For SVM-RBF and RF, we performed a grid search over key hyperparameters (e.g., SVM: C and γ, RF: $n_{\text{estimators}}$ and max_{depth}) using the validation set. For DeepDTA, GCN, GAT and attentive FP, we used the authors' original architectures and conducted hyperparameter searches consistent with their publications or using our validation set. All deep learning baselines were trained under the same hardware and software environment as our MC-GNNVS model.

4.1.3 Evaluation metrics and implementation details.

We employed evaluation metrics commonly used in virtual screening: area under the curve (AUC), enrichment factor (EF1%) (assessing the enrichment capability of active molecules within the top 1% ranking), and area under the precision-recall curve (AUPRC). All deep learning models were implemented using PyTorch and PyTorch Geometric libraries, trained on NVIDIA Tesla V100 GPUs. We employed the AdamW optimiser with an initial learning rate of 0.001 and a batch size of 128. Learning rate decay and early stopping strategies were implemented to prevent overfitting.

4.2 Results and analysis

4.2.1 Comparison of key findings

To visually illustrate the comparison results, we have created Figure 2.

4.2.1.1 Statistical analysis

To rigorously evaluate the significance of performance differences, we conducted paired statistical tests. For the DUD-E benchmark, each of the 102 protein targets constitutes an independent data point (i.e., per-target AUC/EF1%/AUPRC). For MC-GNNVS and every baseline model, we performed 10 independent training runs with different random seeds (affecting weight initialisation, dropout, and data order). The performance metric for a given target and model is reported as the mean ± standard deviation across these 10 runs. Statistical significance between MC-GNNVS and each baseline was assessed using a paired, two-tailed Student's t-test on the 102 pairs of per-target mean scores. A p-value < 0.05 was considered statistically significant. The results confirm that MC-GNNVS's improvements are significant against all

baselines ($p < 0.05$). Table 1 summarises the average performance of all comparison methods on the DUD-E test set.

Figure 2 Performance comparison bar chart (see online version for colours)

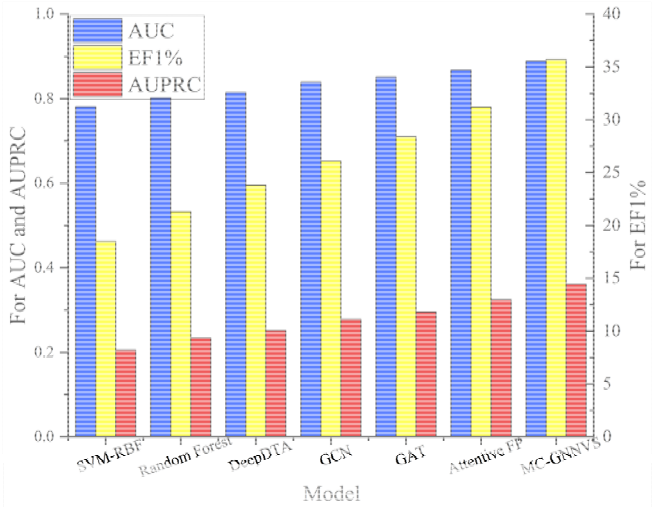
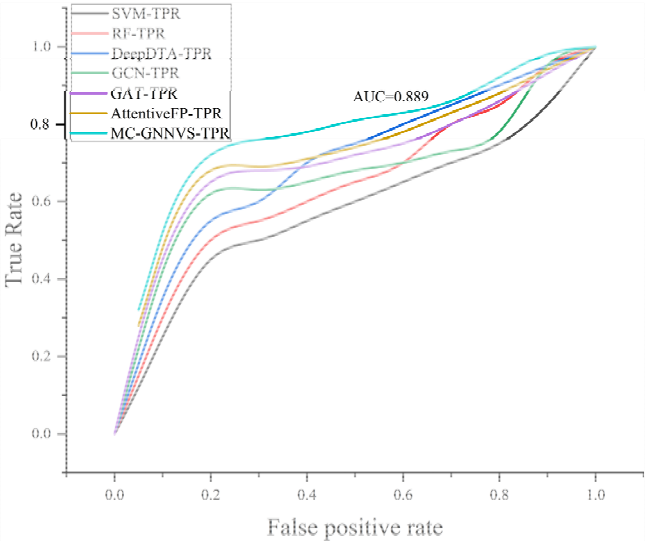


Table 1 Performance comparison of different virtual screening methods on the DUD-E test set (mean ± standard deviation)

Model	*AUC*	*EF1%*	*AUPRC*
SVM-RBF	0.781 ± 0.045	18.5 ± 6.2	0.205 ± 0.051
Random forest	0.802 ± 0.038	21.3 ± 7.1	0.234 ± 0.048
DeepDTA (adapted)	0.815 ± 0.041	23.8 ± 6.5	0.251 ± 0.055
GCN	0.839 ± 0.036	26.1 ± 5.9	0.278 ± 0.049
GAT	0.851 ± 0.033	28.4 ± 5.7	0.295 ± 0.046
Attentive FP	0.868 ± 0.031	31.2 ± 5.3	0.324 ± 0.044
MC-GNNVS (ours)	0.889 ± 0.028	35.7 ± 4.8	0.361 ± 0.040

The results clearly demonstrate that our proposed MC-GNNVS model achieves the best performance across all three metrics. Specifically, compared to the strongest baseline model, attentive FP, MC-GNNVS improves AUC by approximately 2.4% and enhances the more practically significant EF1% metric by over 14%. This outcome demonstrates that by integrating graph structure and sequence information, the model captures richer molecular features, thereby more accurately ranking highly active molecules at the top. Results from two-tailed t-tests on all model pairs indicate that MC-GNNVS exhibits statistically significant performance improvements (p-value < 0.05) relative to all baseline models.

Figure 3 Receiver operating characteristic curve comparison chart (see online version for colours)



4.2.2 *Melting experiment*

To understand the contributions of each component in MC-GNNVS, we designed ablation experiments, with the results shown in Table 2.

Table 2 Dissolution experiment results (average AUC on the ChEMBL kinase target dataset)

Model variants	*Description*	*AUC*
MC-GNNVS (full)	Complete model	0.882
w/o sequence channel	Remove SMILES sequence channel	0.865
w/o attention pooling	Replace attention pooling with global average pooling	0.854
w/o edge features	Removing edge features in GAT	0.848
GAT only	Using only the GAT channel, with no enhancements whatsoever.	0.851

Table 3 Computational efficiency comparison for screening one million compounds

Method	*Hardware*	*Avg. throughput (molecules/sec)*	*Total time*	*Preprocessing time*	*Inference time*
Molecular docking (reference)	CPU cluster	~0.1	~115 days	N/A	N/A
Random forest (ECFP4)	Single CPU (Xeon)	~1,200	~14 min	~2 min	~12 min
MC-GNNVS (ours)	Single GPU (V100)	~1,100	~15 min	~3 min	~12 min

Note: Docking time is an estimate based on typical literature values. RF and MC-GNNVS times are measured end-to-end on our test system. The comparable throughput of MC-GNNVS, while delivering superior accuracy (AUC 0.889 vs. 0.802), highlights its favourable accuracy-efficiency trade-off.

The ablation results demonstrate that each component of the model is essential. Removing the SMILES sequence channel leads to a significant performance drop, confirming the necessity of multimodal fusion. Sequence information can compensate for certain sequence-dependent syntactic information that may be lost in the graph structure. Both performance loss is also observable in the case of replacing attention pooling by average pooling, which means that the performance of letting the model learn which atoms are important in global tasks is better than that by mere averaging. Intuitively, this mimics a medicinal chemist's focus on key functional groups or pharmacophores when assessing activity, rather than treating all atoms equally. Moreover, the removal of edge features also had an undesirable impact on performance, which further highlights the importance of such information as the type of bonds in the correct representation of molecular structures.

4.2.3 Efficiency analysis and case studies

4.2.3.1 Computational efficiency benchmark

We evaluated the practical throughput of MC-GNNVS on a standard screening task. All timing experiments were conducted on an identical hardware platform: a single NVIDIA Tesla V100 GPU (32 GB) with an Intel Xeon CPU. We measured the end-to-end time to screen one million molecules from the ZINC15 database, including file I/O, SMILES parsing, graph generation, model inference, and result saving. For a fair comparison, we also measured the throughput of the random forest (RF) baseline – the strongest traditional ML model in our study – on the same CPU using its optimised scikit-learn implementation. As a reference point, we cite the typical timescale for high-throughput molecular docking on CPU clusters. The results, summarised in Table 3, demonstrate that MC-GNNVS achieves GPU-accelerated screening speed comparable to CPU-based RF, but with significantly higher predictive accuracy, highlighting an excellent accuracy-efficiency trade-off.

In addition to objective metrics of performance, it is critical to appreciate the decision-making logic of the model and earn the trust of medicinal chemists, as well as provide future optimisation. We applied the visualisation tools to critically examine the internal attention weights in the model to investigate the chemical rationality of the prediction the model makes. The model was effective in assigning the highest weight of the attention to a known hinge-binding fragment in the molecule in a representative case with validated results of kinase inhibitor. This fragment is biochemically proved to be the one that contributes to creating essential interactions of dihydrogen bonds with the target protein, acting as a core pharmacophore as needed to preserve high activity. Our model was able to find this important substructure and target it without any prior chemical background knowledge being explicitly provided. The fact that this finding greatly confirms the biophysical nature of the decision-making mechanism in the model, as opposed to data-fitting, also underscores that the model application has enormous possibilities in optimisation of lead compounds. The decryption of the model of points of interest would allow chemists to more reliably rationally tweeze and adjust molecules, including maintaining and improving this high-attention fragment. This demonstrates that the model's interpretability is not merely a supplementary feature but a core strength, effectively bridging AI-driven prediction with human chemical intuition to guide actionable molecular optimisation.

5 Research summary

This paper was able to design and optimally test a drug molecule virtual screening system founded on a multi-channel graph neural network (MC-GNNVS). A lot of experimentation on public benchmark datasets, including DUD-E and ChEMBL, has shown that this system can show a substantial and statistically significant increase in performance relative to traditional machine learners (e.g., random forests) and contemporary deep learning models (e.g., standard GCN, GAT and attentive FP). In particular, the system raised the area under the curve (AUC) of the screening to 0.889 and the EF1% score to 35.7 – an important measure of early enrichment capability. This illustrates the capability of the system to be more accurate and fast in the detection of potential active molecules of large compound libraries, aptly handling the high failure rate in the drug discovery first phases.

Any theoretical contribution of this work is mainly depicted in two aspects. First of all, we present and justify a powerful paradigm of multimodal fusion of molecular information that combines the topological data of the molecular graph structure with the syntactic data of molecular sequences in the form of SMILES. The experiments of ablation prove that the absence of either of the information channels results in a substantial drop in model performance (e.g., a graph structure alone yields an AUC of 0.865), which proves the complementary character of the various memory representations. The combination of them makes the model more comprehensive and discriminative in its feature representation. Second, the system combines attention systems and residual connections. This does not just allow the model to concentrate on key sub-structural elements of chemistry, so greater interpretability, but it also provides stability during the deep training of network systems, which serve as on-hand technical reference when building up more complex molecular structures.

On the practical level, the given study offers an effective and trustworthy solution to computer-aided drug discovery. The MC-GNNVS system is highly balanced (at around 15 minutes to complete a screening task on a million-molecule library) and can be used in large-scale virtual screening due to its practical viability. What is even more important is that the major pharmacophores that are disclosed in regard to attention weight visualisation are in line with the previous understanding of medicinal chemists. This goes a long way in gaining confidence among the

researchers in the artificial intelligence models and give them reasonable directions and in depth chemical knowledge to further optimisation of the lead compound.

5.1 Discussion and future perspectives

While MC-GNNVS demonstrates superior performance, we acknowledge several limitations that present avenues for future work. First, our model's current architecture is primarily ligand-centric. A critical next step is to evolve it into a full drug-target interaction prediction system. This can be achieved by integrating target protein information (e.g., as 1D sequences via transformers or 3D structures via geometric deep learning) and employing cross-attention mechanisms to model the physical interaction between the molecule and the target protein. Second, the model's performance is contingent upon the availability of high-quality, labelled activity data, which is often scarce and expensive to obtain. To mitigate this, we plan to explore self-supervised pre-training strategies on large unlabeled molecular databases (e.g., ZINC, PubChem). By pre-training on tasks like masked atom prediction or context prediction, the model can learn rich, generalised molecular representations, which can then be fine-tuned on smaller, labelled datasets, potentially enhancing performance and data efficiency. Third, to further de-risk the drug discovery pipeline, we intend to extend our framework into a multi-task learning system that simultaneously predicts primary activity alongside key absorption, distribution, metabolism, excretion, and toxicity (ADMET) properties. Developing a model that can balance potency with favourable pharmacokinetics and safety profiles at the virtual screening stage would be a significant advancement. Finally, while attention mechanisms provide some interpretability, the explanations are still post-hoc. Future work will focus on developing inherently interpretable architectures or using generative models to provide actionable suggestions for molecular optimisation, moving beyond highlighting important substructures towards proposing specific chemical modifications.

Declarations

The author declares that she has no conflicts of interest.

References

Albarillo, C.V. and Fernandez, P.L. (2023) 'Investigating the effect of BD-CRAFT to text detection algorithms', *International Journal of Artificial Intelligence & Applications*, Vol. 14, No. 1, p.16.

Ann, V. and William, R. (2025) 'Multidisciplinary perspectives on cumulative impact assessment for vulnerable communities: expert elicitation using a Delphi method', *Integrated Environmental Assessment and Management*, Vol. 2, No. 3, p.12.

Esakia, T., Sharikadze, N., Janadze, K., Kekelia, E. and Abuladze, M. (2025) 'Patterns and barriers of second-line therapy selection in HR+/HER2– metastatic breast cancer among medical oncologists in a developing country (Georgia)', *Journal of Clinical Oncology*, Vol. 43, No. 16, p.13067.

Estrela, V.V. and Hemanth, J. (2015) 'International journal of information and communication technology, special issue on medical informatics and technology', *International Journal of Information and Communication Technology*, Vol. 1, No. 2, p.6.

Ferreo, D., Kirk, M., Serrano, M. and Sainz-Aja, J. (2022) 'Assessment of the generalization ability of the ASTM E900-15 embrittlement trend curve by means of Monte Carlo cross-validation', *Metals*, Vol. 5, No. 3, p.11.

Gilmer, J., Schoenholz, S.S., Riley, P.F., Vinyals, O. and Dahl, G.E. (2017) 'Neural message passing for quantum chemistry', *Proceedings of the 34th International Conference on Machine Learning (ICML)*, Vol. 1, No. 1, pp.1263–1272.

Hemanth, D.J. and Estrela, V.V. (2013) 'Call for papers for the *International Journal of Information and Communication Technology* special issue on: 'Medical informatics – a practical perspective'', *IEEE Sensors Journal*, Vol. 3, No. 4, p.2.

Hosseini, S.M., Zarei, T. and Pierantozzi, M. (2024) 'Modeling equilibrium and non-equilibrium thermophysical properties of liquid lubricants using semi-empirical approaches and neural network', *Journal of Non-Equilibrium Thermodynamics*, Vol. 49, No. 3, p.8.

Ishida, M.B.T., emailprotected, Emailprotected, E., Ishida, T., M.B.T. (2023) 'Comparing supervised learning and rigorous approach for predicting protein stability upon point mutations in difficult targets', *Journal of Chemical Information and Modeling*, Vol. 63, No. 21, pp.6778–6788.

Johannes, R. (2022) 'A neural-symbolic approach for explanation generation based on sub-concept detection: an application of metric learning for low-time-budget labeling', *KI – Künstliche Intelligenz*, Vol. 36, Nos. 3–4, pp.225–235.

Kim, H., Lee, K., Kim, J.H. and Kim, W.Y. (2024) 'Deep learning-based chemical similarity for accelerated organic light-emitting diode materials discovery', *Journal of Chemical Information and Modeling*, Vol. 64, No. 3, p.13.

Lam, S., Liu, Y., Broers, M., Vos, J.V.D., Frasincar, F., Boekestijn, D. and Knaap, F.V.D. (2025) 'Local interpretation of deep learning models for aspect-based sentiment analysis', *Engineering Applications of Artificial Intelligence*, Vol. 143, No. 2, p.23.

Li, Z., Ding, N., Liang, C., Cao, S., Zhai, M., Huang, R., Zhang, Z. and Hu, B. (2025) 'A semi-supervised framework fusing multiple information for knowledge graph entity alignment', *Expert Systems with Application*, Vol. 1, No. 2, p.259.

Matsumura, Y., Tabata, K. and Komatsuzaki, T. (2025) 'Comparative analysis of reinforcement learning algorithms for finding reaction pathways: insights from a large benchmark data set', *Journal of Chemical Theory and Computation*, Vol. 21, No. 7, pp.3523–3535.

Mysinger, M.M., Carchia, M., Irwin, J.J. and Shoichet, B.K. (2012) 'Directory of useful decoys, enhanced (DUD-E): better ligands and decoys for better benchmarking', *Journal of Medicinal Chemistry*, Vol. 55, No. 14, p.6582.

Truchon, J.F.O. and Bayly, C.I. (2007) 'Evaluating virtual screening methods:? good and bad metrics for the 'early recognition' problem', *Journal of Chemical Information & Modeling*, Vol. 47, No. 2, pp.488–508.

Venkatesan, H., Mahesh, J.S. and Venkataraman, S. (2025) 'Decoding hepatitis B virus mutations that impact host-virus interactions and therapeutics', *Journal of Biosciences*, Vol. 50, No. 3, pp.1–12.

Wang, M., Wu, Z., Wang, J., Weng, G., Kang, Y., Pan, P., Li, D., Deng, Y., Yao, X. and Bing, Z. (2024) 'Genetic algorithm-based receptor ligand: a genetic algorithm-guided generative model to boost the novelty and drug-likeness of molecules in a sampling chemical space', *Journal of Chemical Information and Modeling*, Vol. 64, No. 4, p.16.

Wu, Z., Ramsundar, B., Feinberg, E.N., Gomes, J., Geniesse, C., Pappu, A.S., Leswing, K. and Pande, V. (2018) 'MoleculeNet: a benchmark for molecular machine learning', *Chemical Science*, Vol. 9, No. 2, pp.513–530.

Yang, X., Wang, Y., Byrne, R., Schneider, G. and Yang, S. (2019) 'Concepts of artificial intelligence for computer-assisted drug discovery', *Chemical Reviews*, Vol. 119, No. 18, pp.10520–10594.

Zhu, J., Min, A., Lerner, S., Phair, J., Vouyouka, A., Smolock, C., Marin, M., Faries, P., Han, D. and Rao, A. (2024) 'Development of a surgical skills course for preclinical medical students', *JVS – Vascular Insights*, Vol. 2, No. 3, p.10.