Computational and Structural Biotechnology Journal 27 (2025) 4633–4644

Contents lists available at ScienceDirect

Computational and Structural Biotechnology Journal

journal homepage: www.elsevier.com/locate/csbj

Research article

Predicting PROTAC off-target effects via warhead involvement levels in drug–target interactions using graph attention neural networks

Yutong Hu [a,b], Kieran Didi [c], Adam P. Cribbs [d], Jianfeng Sun [a,d,*]

[a] Thrust of Bioscience and Biomedical Engineering, Hong Kong University of Science and Technology (Guangzhou), Guangzhou 511400, China
[b] Department of Data Science and Technology, Hong Kong University of Science and Technology, Hong Kong, SAR 999077, China
[c] Department of Computer Science, University of Oxford, Oxford OX1 3QG, UK
[d] Botnar Research Centre, University of Oxford, Oxford OX3 7LD, UK

ARTICLE INFO

Keywords:
PROTACs
Degradation
Off-target effects
Graph attention networks
Deep learning

ABSTRACT

Proteolysis-targeting chimeras (PROTACs) represent an emerging modality for targeted protein degradation with broad therapeutic potential. However, the risk of off-target protein degradation remains a major concern in the development of PROTAC-based therapeutics. Here, we present SENTINEL, a graph-based deep learning framework that predicts the off-target propensity of PROTAC warheads based on their involvement levels in drug–target interactions as determined from established databases and the literature. By encoding warheads as molecular graphs using path-augmented graph transformer networks (PAGTNs), we show that graph attention-based neural networks (GATs) achieve accurate modelling of binding count-based off-target effects with an area under the ROC curve (AUC) of 0.9600 and an F1-score of 0.6983, outperforming classical machine learning algorithms such as random forests (AUC=0.840, F1-score=0.2778). SENTINEL provides a scalable strategy to prioritise lower-risk warheads in a low-data setting, supporting early-stage evaluation of PROTAC off-target risk. Results should be interpreted with the dataset size in mind and will benefit from larger external validation.

1. Introduction

Proteolysis-targeting chimeras (PROTACs) enabled targeted protein degradation by recruiting E3 ubiquitin ligases (E3s) to induce proteasomal clearance [1]. Unlike traditional inhibitors that block protein functions through sustained occupancy [2], PROTAC-catalysed proteasomal degradation can lead to the complete removal of both enzymatic and scaffolding proteins [3,4]. This event-driven mechanism can expand therapeutic options to the previously "undruggable" proteins of interest (POIs) and may allow durable pharmacological effects at lower doses [5–7]. These features of PROTACs have motivated widespread efforts to optimise therapeutic agents across diverse diseases [8,9], making the systematic assessment of binding specificity and off-target risks an important step in drug development [10,11].

PROTAC off-target effects can arise via multiple mechanisms, including non-specific E3 ligase recruitment and POI warhead-mediated binding with unintended proteins [12]. The E3 ligase-recruiting moiety may inadvertently localise ubiquitination machinery to proteins proximal to the intended target, leading to unintended ubiquitination and degradation. For example, pomalidomide, one of immunomodulatory imide drugs (IMiDs) [13] and also a commonly used E3-recruiting ligand in PROTACs [14,15], can induce degradation of multiple zinc-finger (ZF) proteins involved in diverse cellular processes [16], e.g., the ZFP91 protein for transcriptional activation [17]. In addition, the POI warhead of a PROTAC may interact with off-target proteins exhibiting structural or physicochemical similarity to the intended target, leading to their unintended degradation. For example, studies suggest that cereblon (CRBN)-based PROTACs can degrade multiple bromodomain (BRD) proteins with which the POI warhead moiety binds [18–20]. Moreover, POI warheads may also engage off-target proteins that lack obvious structural or sequence similarity to the intended POI. For example, one study reported down-regulation of the breakpoint cluster region (BCR) protein, likely due to its unintended binding of the estrogen-related receptor-α (ERRα) ligand in a PROTAC that was commonly used to degrade ERRα the proteins [21]. Despite growing recognition of PROTAC off-target effects and advances in proteomics-based profilling, as noted in [22], scalable computational tools to proactively assessing the off-target potential of PROTACs remain

* Corresponding author at: Thrust of Bioscience and Biomedical Engineering, Hong Kong University of Science and Technology (Guangzhou), Guangzhou 511400, China.
E-mail addresses: jianfeng.sun@ndorms.ox.ac.uk, jianfengsun@hkust-gz.edu.cn (J. Sun).

https://doi.org/10.1016/j.csbj.2025.10.028
Received 1 August 2025; Received in revised form 15 October 2025; Accepted 16 October 2025
Available online 19 October 2025

limited. Here, we focus specifically on POI-warhead-mediated off-target effect prediction, leveraging the availability of numerous warheads catalogued in established drug-target interaction (DTI) databases.

Advances in PROTAC databases have enabled computational strategies for structural modelling, feature extraction, and intrinsic property prediction [23,24]. Most existing computational approaches emphasise PROTAC design [25–27], whereas comparatively fewer address degradation potential [28,29] or activity prediction [30]. These models target the question of how effectively a PROTAC degrades its intended target. The emergence of large language models (LLMs) has further accelerated research generative design of PROTAC molecules [31]. However, yet another challenge is to predict the potential of a PROTAC for off-target effects. This requires assessing the warhead interaction profile across the entire proteome. To date, computational approaches for predicting the off-target effects associated with PROTACs remain limited.

Here, we propose SENTINEL, a computational approach for predicting PROTAC off-target effects arising from POI-warhead-mediated interactions with unintended proteins. To this end, we systematically retrieved, from curated DTI databases, small molecules annotated as POI-warhead moieties in known PROTACs and compiled, for each warhead, its curated set of protein interaction partners. We utilised the number of unique target proteins associated with a given POI warhead as a proxy for its off-target propensity. For every warhead identified by the PubChem compound identifier (CID), we deduplicate DTI records at the CID-by-target level and count unique targets. This count is then classified into three classes as an interpretable surrogate for off-target propensity. For predictive modelling, we encode POI-warhead molecular graphs with a path-augmented graph transformer network (PAGTN) [32], and train a graph attention-based network (GAT) [33] on these representations to improve performance of off-target effect prediction. By classifying POI warheads into risk levels based on the number of associated gene targets, this approach aims to

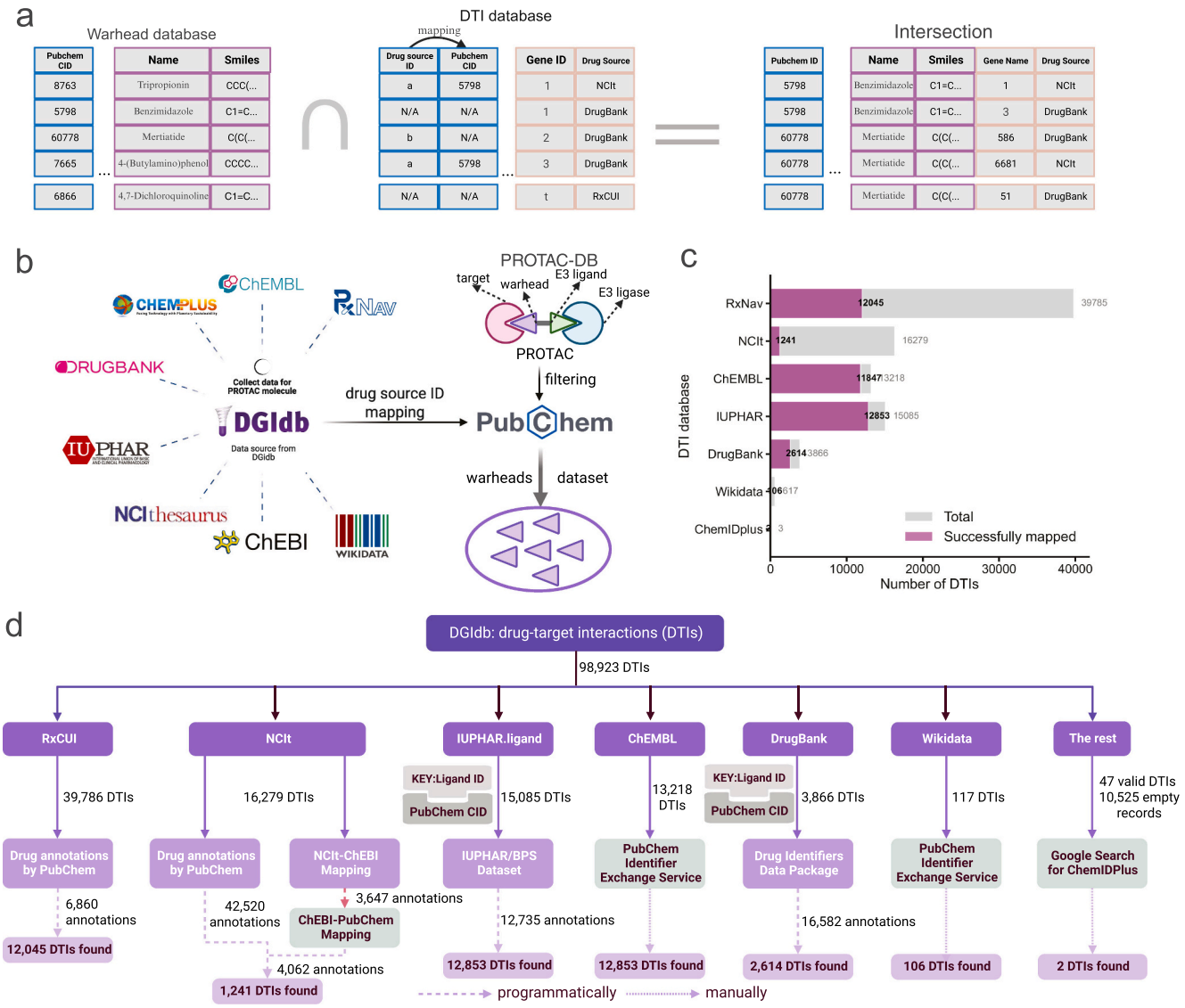

Fig. 1. Workflow of linking PROTAC warheads to experimentally verified drug–target interactions (DTIs). (a). Schematic of the identifier matching process between PROTAC warheads and DTI entries. Warheads were matched to known DTIs through PubChem CIDs to identify overlapping compounds with available target annotations. (b). Overview of PROTAC and DTI sources. Warheads were retrieved from PROTAC-DB and linked to DTIs aggregated from multiple public databases via DGIdb. Identifier mapping was used to ensure consistent compound representation. (c). Summary of DTI coverage across data sources. The number of total and successfully mapped DTIs varies across databases depending on annotation availability. (d). Specific mapping workflow on individual databases. PubChem annotations, external mapping files, and manual curation steps were combined to reconcile drug identifiers to PubChem CIDs. Only DTI entries with successfully matched warheads were retained for downstream analysis.

provide an interpretable metric for early-stage evaluation of off-target risks and potential cytotoxicity of PROTACs, thereby contributing to their safer development.

2. Results

2.1. Off-target effect prediction using DTI counts

To model POI warhead-induced off-target effects of PROTACs, we established a data integration pipeline that reconciles warhead annotations from PROTAC-DB (Database) with known drug–target interactions (DTIs) from eight publicly available databases, such as DrugBank and ChEBI (Fig. 1**a**). We began by extracting and filtering PROTACs, retaining 479 warheads with valid PubChem compound identifiers (CIDs). Simultaneously, we curated 98,923 DTIs from eight major source databases (e.g., ChEMBL, DrugBank, IUPHAR, NCIt), each of which provides drug identifiers in source-specific formats (Fig. 1**b**). By combining automated matching of warhead identifiers with PubChem CIDs (e.g., using the PubChem Identifier Exchange Service) and manual curation for sources with limited coverage, we obtained more than 40,000 DTI candidates with explicitly tractable identifiers (Figs. 1**c and** 1**d**). Then, we intersected the curated DTI dataset with PROTAC-DB warheads based on PubChem CIDs. This step yielded 172 warheads that were both structurally annotated in the PROTAC database and functionally linked to known protein targets through validated DTIs. This set of warheads was used for target quantification, off-target labeling, and machine learning classification. Detailed introduction to identifier mapping and data pre-processing can be found in section *Methods*.

2.2. GAT for off-target effect prediction

We constructed a graph attention network (GAT)-based architecture to model the off-target effect potential of PROTAC warheads (Fig. 2). The final model (SENTINEL) uses graph-structured inputs only. Each molecule is first represented as a graph, where atoms and bonds are encoded as 94-dimensional node features and 42-dimensional edge features, respectively. The molecular graph is processed through two successive GAT convolutional layers, followed by rectified linear unit (ReLU) activation and a global mean pooling operation to generate a fixed-size, graph-level embedding (Fig. 2**a**). The core mechanism of the GAT layers relies on learnable attention coefficients to adaptively weigh neighboring nodes and edges during feature aggregation. As illustrated in Fig. 2**b**, attention weights are computed separately with trainable parameters, followed by a leaky rectified linear unit (LeakyReLU) activation to generate attention coefficients for warheads. These attention coefficients enable the model to focus selectively on the informative parts of warheads molecular graphs. The GAT functions as a classifier and outputs three risk tiers defined by the target-count threshold (see Section 2.3 for details).

For ablation study only, we added a parallel branch that incorporated nine handcrafted physicochemical descriptors, such as molecular weight, hydrogen bond donor/acceptor counts and rotatable bonds, into a two-layer fully-connected neural network (Fig. 2**c**). These features were first transformed by a hidden layer with 64 neurons, followed by an identical layer with 64 neurons, with ReLU activations applied at each stage. This produces a complementary embedding that encodes chemically interpretable information beyond what is captured by graph structure alone. We concatenated the graph-based and descriptor-based embeddings and passed the joint representation through a dropout layer

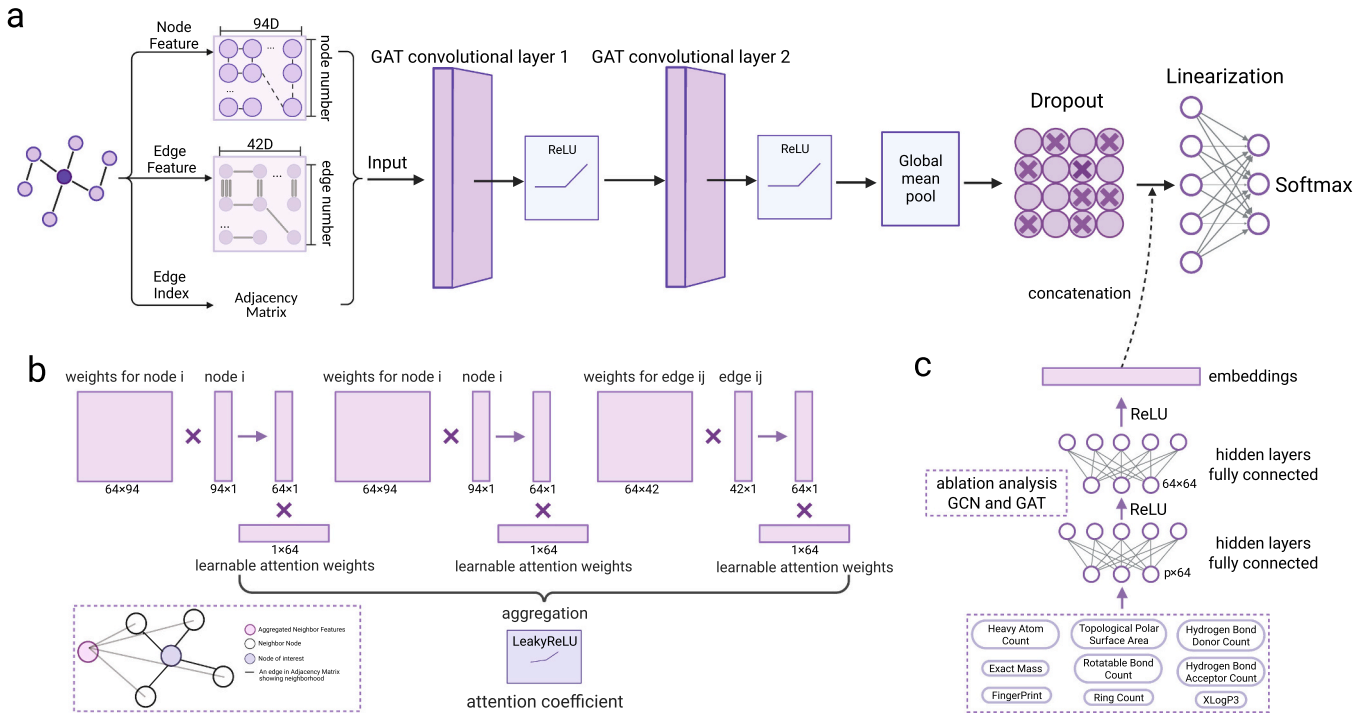

Fig. 2. Overview of the SENTINEL architecture. (a). Architecture of the graph attention network (GAT). The model takes molecular graph representations as input, with node features (94-dimentional, 94D) and edge features (42-dimentional, 42D) representing atoms and bonds, respectively. These are processed by two stacked GAT convolutional layers, each followed by a ReLU activation. Global mean pooling is applied to obtain molecular embeddings of contracted size, which are subsequently passed through a dropout layer and fully connected layers to produce 3-class softmax outputs. (b). Attention mechanism introduced to the GCN. Separate attention coefficients are learned for node and edge features using trainable weight matrices and LeakyReLU activation. This allows adaptive aggregation of local graph information by weighting neighboring nodes and edges. (c). Handcrafted feature embedding module used exclusively in the ablation study. A total of nine physicochemical descriptors (e.g., hydrogen bond counts, exact mass, XLogP3, fingerprints) are embedded using a two-layer fully connected neural network with ReLU activation. The resulting embeddings are concatenated with graph-based embeddings and fed into a single-layer fully-connected neural network to inform warhead-induced off-target effects of PROTACs.

and a final softmax classifier to predict the off-target potential label of an input warhead. This design allowed for evaluating the relative contribution of graph-derived versus feature-derived signals under different experimental settings (see section *Methods* for details).

2.3. Optimising labeling strategies to inform warhead-induced off-target effects

We performed a comparative evaluation to determine a threshold-

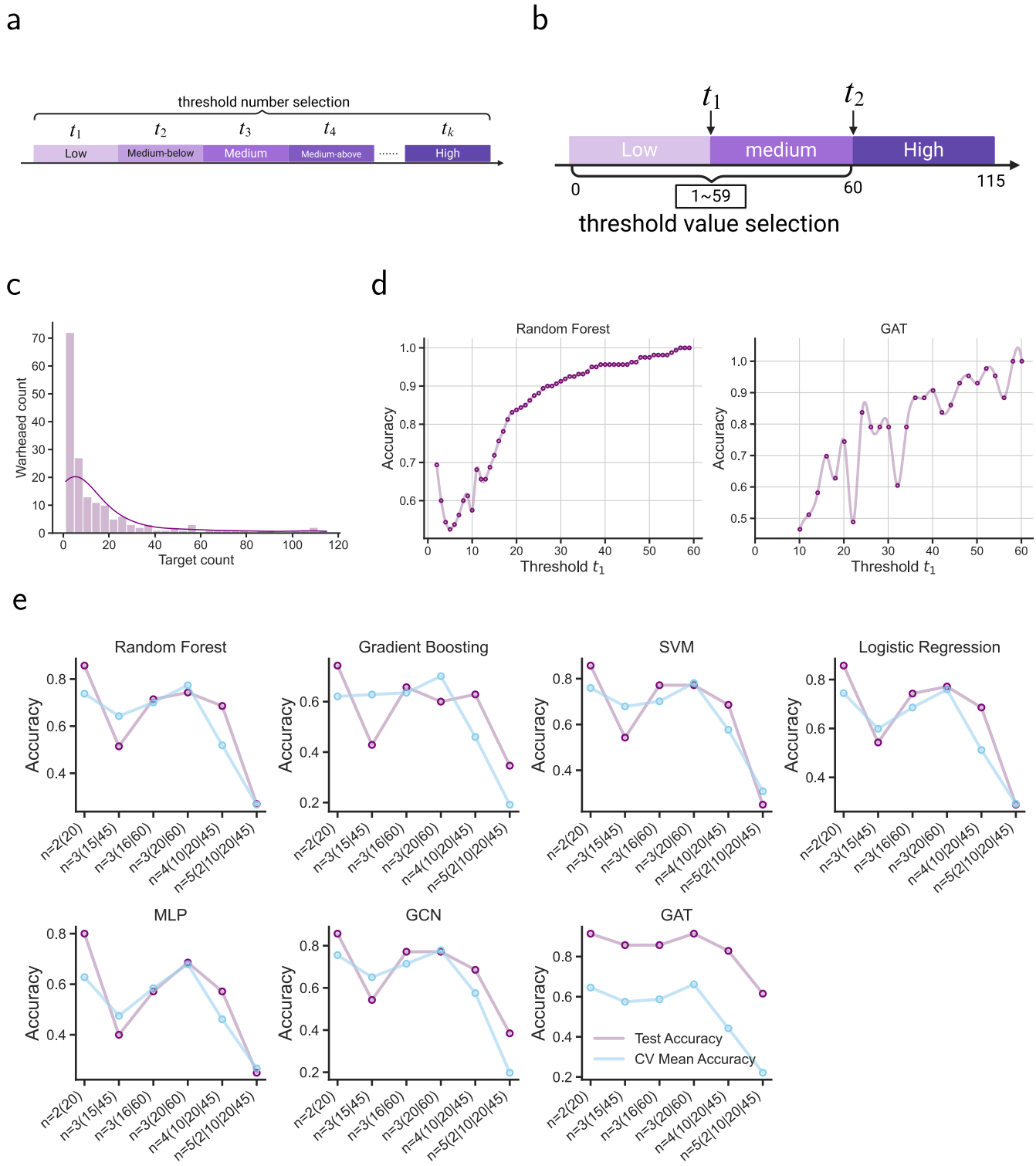

Fig. 3. Analysis of threshold-based labeling strategies for gauging warhead off-target effects. (a). Schematic showing the definition of label categories using different numbers of thresholds. Increasing or decreasing the number of thresholds allows the potential for finer-grained classification of off-target potential levels. (b). Strategy for selecting cutoff values in a 3-label classification scheme based on 2 thresholds. The upper threshold t_2 was fixed at 60 to capture highly promiscuous warheads, while the lower threshold t_1 was optimised based on model performance. (c). Distribution of warheads by their target counts. There is a skewed profile where most warheads interact with a small number of targets. (d). Performance of random forest and GAT models across various t_1 values. (e). Comparison of test and cross-validation accuracy across different labeling strategies using various machine learning models.

based labeling strategy to characterise the propensity of warheads to induce off-target effects. An effective labeling strategy depends on two key factors: *i*) the number of thresholds selected to define discrete categories, and *ii*) how appropriately these thresholds partition the warheads according to the number of associated DTIs. We evaluate both the number and placement of thresholds based on their influence on the classification performance of machine learning models. Lastly, to maximise the generalisation ability of the models, we ensure that the similarity between training and testing small molecules remains within a threshold reported in the literature (See *Methods*) [34].

2.3.1. Evaluation of multi-label strategies for warhead stratification

We began by investigating how the number of classification labels affects model performance. These labels were derived by applying the following 4 thresholding schemes to the number of targets, POIs that each PROTAC warhead is designed to bind or potentially degrade, associated with each warhead.

$$\begin{cases} 2labels, \& t_1 = 20 \\ 3labels, \& t_1 = 20 \quad , \quad t_2 = 60 \\ 4labels, \& t_1 = 10 \quad , \quad t_2 = 20 \quad , t_3 = 45 \\ 5labels, \& t_1 = 2 \quad , \quad t_2 = 10 \quad , t_3 = 20, t_4 = 45 \end{cases}$$

In the 2-label setup, warheads are stratified into low and high off-target potential. The 3-label setup stratifies warheads into low, medium, and high off-target potential. The 4-label setup stratifies warheads into low, medium-below, medium, and high off-target potential. The 5-label setup stratifies warheads into low, medium-below, medium, medium-above, and high off-target potential. We found that graph neural network models generally outperformed other machine learning methods across different label configurations. Particularly, the GAT model achieved higher accuracy (Fig. 3**a**), precision (Fig. 3**b**), and recall (Fig. 3**c**) likely due to its ability to capture complex graph-structured relationships via attention mechanisms. Although the 2-label configuration yielded the highest classification performance in terms of standard machine learning metrics, we observed that a single threshold fails to capture the full variability in the distribution of DTI counts across warheads (Fig. 3**d**). The data is skewed, with a large number of warheads having few DTIs. This non-uniform distribution suggests that a binary classification may oversimplify the underlying biological heterogeneity. Therefore, we adopted a 3-label strategy, which strikes a pragmatic balance between model performance and data distribution, providing a m representation of warhead off-target potential.

2.3.2. Selection of cutoff values for warhead stratification

To refine the threshold selection for 3-label classification, we aimed to identify optimal cutoff points for t_1 and t_2 that not only align with the underlying distribution of DTI counts but also support relatively high predictive performance. Given that the distribution of DTI counts spans a wide range between 1 and 110, we empirically fixed the upper threshold t_2 at 60 to account for the small proportion of warheads yet with high off-target potential, and focused on determining a suitable lower threshold t_1 to separate the low and medium classes (Fig. 3**e**).

To this end, we trained the machine learning models and evaluated the performance at each $t_1 \in \{1, 2, \ldots, 59\}$. For each candidate threshold, warheads were binned into two initial classes 0 for $[0, t_1]$ and 1 for $(t_1, 59]$. Warheads with a DTI count no smaller than 60 were treated as class 2 in the final multi-class setting. A Random Forest model was trained and evaluated on a validation subset (20 %) of the 138 warheads in the training set. Mean cross-validation accuracy was plotted as a function of t_1 (Fig. 3**f**). The results revealed a sharp increase in the light of accuracy from about 0.6 at $t_1 = 10$ to about 0.7 at $t_1 = 20$, followed by a plateau near 0.9 at $t_1 \geq 30$. Given these significant performance gains observed within the threshold range between 10 and 30, and a relatively noticeable fluctuation in the distribution of DTI counts in this interval, we selected $t_1 = 20$ for low and medium off-target potential as a compromise that mitigates class imbalance while preserving classification accuracy. In addition, we observed a similar changing trend in terms of the evaluation metrics using the GAT model, schematically supporting this choice (Fig. 3**g**). Finally, we retrained the GAT model on the complete set of warheads stratified with $t_1 = 20$ and $t_2 = 60$.

2.4. Performance comparison between models

We evaluated model performance in predicting warhead-induced off-target effects using a comprehensive suite of evaluation metrics (Fig. 4 and Supplementary Tables 1–11). By comparing the GAT model with the graph convolutional network (GCN model), we found that the attention mechanism appears to reinforce the ability of graph neural networks to predict off-target effects, as evidenced by most of the evaluation metrics. Particularly, on the test dataset, by plotting the receiver operating characteristic (ROC) curve, we found that the GAT model (seed 2) achieved an area under the curve (AUC) of 0.960 (Supplementary Figure 1), along with an accuracy of 0.8857, a precision value of 0.8821, an F1-score of 0.6983, and a specificity value of 0.8345. To ensure result stability and account for stochastic variation during training, we trained the GAT model with two random seeds. The GAT (seed 1) model also performed comparably well, with slightly lower but still top-ranking scores at each metric. GCN followed with a test accuracy of 0.7714, F1-score of 0.4355, and specificity of 0.7500, which placed it ahead of all other machine learning methods.

In contrast, during cross-validation, traditional machine learning models showed better performance than graph neural networks in several metrics using the Random split (see *Methods*). We found that the support vector machine (SVM) achieved the highest cross-validation accuracy (0.7965), Jaccard index (0.2964), and ranked the second best in F1-score (0.3269) and specificity (0.6667), outperforming both GCN (0.7783 in accuracy and 0.2594 in Jaccard index) and GAT models (0.7034/0.6965 in accuracy and 0.2655 in Jaccard index). Besides, the random forest model performed the best in F1-score (0.3249) and recall (0.3627) among all methods except GAT (seed 1), and achieved a strong accuracy of 0.7892, only slightly behind SVM. Moreover, logistic regression and gradient boosting also remained competitive, with the former achieving an accuracy value of 0.782 and an F1-score of 0.3238 and the latter achieving the highest precision (0.3158) among all models. This performance gap is particularly pronounced for the GAT models, whose cross-validation accuracies drop to 0.7034 (seed 1) and 0.6965 (seed 2), significantly lower than those of SVM (0.7965) and the random forest model (0.7892). We speculate that this discrepancy likely reflects limited size of the training dataset, which makes deep architectures unable to extract transferable patterns across folds, leading to the degraded performance. To address potential overfitting and split-dependent variance, we additionally report cluster-aware group cross-validation (Butina) alongside Random splits to reduce chemotype leakage (Supplementary Tables 12-20).

Furthermore, attention-based interpretability analysis confirmed that SENTINEL learns chemically meaningful substructures, assigning higher relative attention to functional motifs (e.g., amides and halogenated aromatics) (Supplementary Table 21 and Supplementary Figs. 5 and 6). This property provides intrinsic interpretability without reliance on external descriptors or post-hoc surrogate models.

2.4.1. Family-level off-target effect evaluation

We next evaluated off-target effect prediction on held-out molecules specific to target families. For each family, we compared the distribution of predicted vs. true risk classes (within-family vs. out-of-family bins) on testing data (unseen). As shown in Supplementary Fig. 7, we found that although several protein families are targeted by only 1-2 warheads, those targeted by $\geq$3 compounds showed strong agreement between predictions and ground-truth values, indicating the positive transfer of learned chemistry at the family level. Full counts and accuracies are provided in Supplementary Table S22.

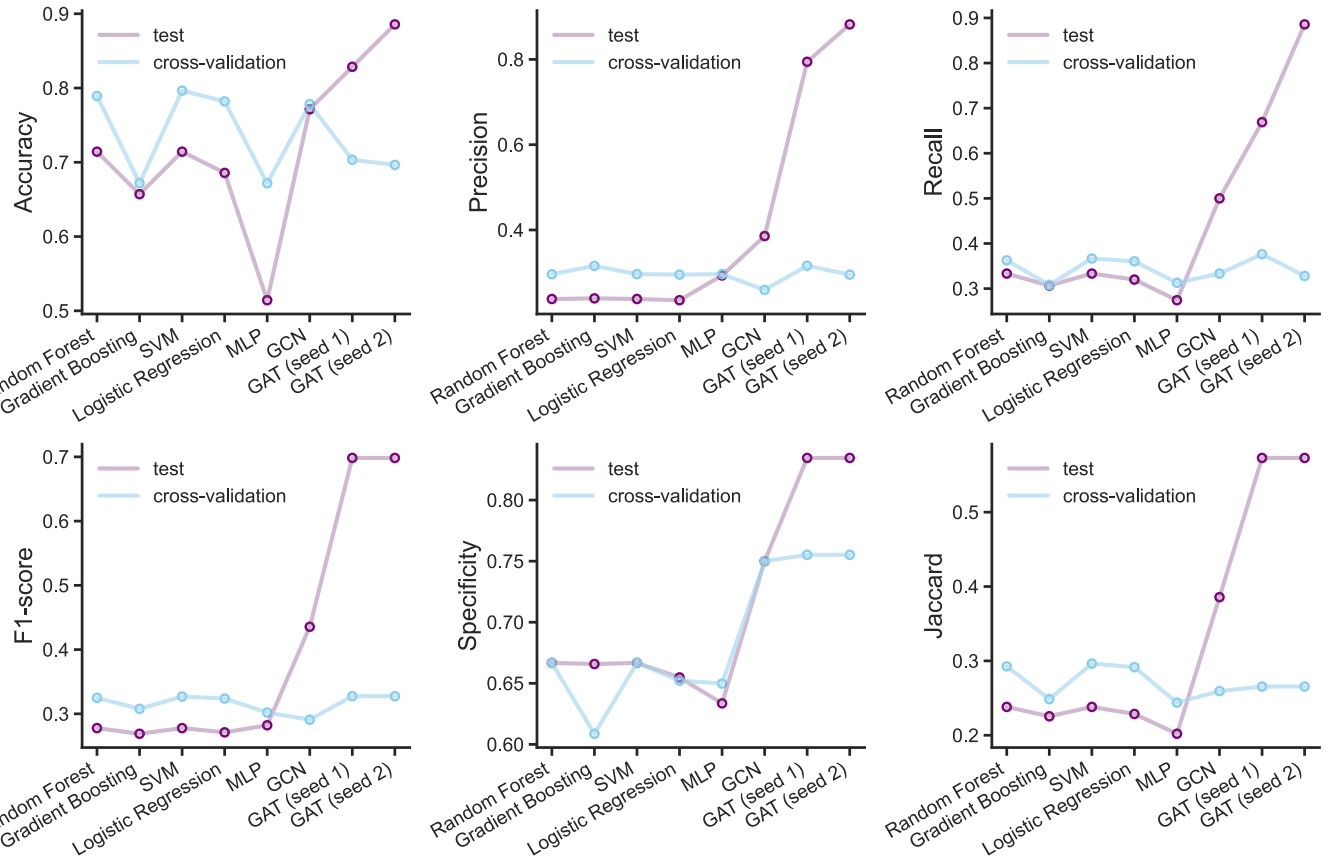

Fig. 4. Performance comparison of machine learning models in predicting PROTAC warhead off-target potential in terms of accuracy, precision, recall, F1-score, specificity, and Jaccard index. For each model, both test set performance colored in purple and cross-validation performance colored in blue are shown.

2.5. Feature importance analysis

Feature ablation analysis is an effective approach to assess the contribution of individual features or feature groups to the optimisation of model interpretation [35]. Both graph-based and non-graph-based features were used. Among the non-graph-based inputs, molecular fingerprints present a challenge due to their high dimensionality, with different dimensional settings potentially influencing ablation outcomes. To address this, we evaluated which fingerprint dimensionality was best suited for machine learning models. Based on this assessment, a standardised feature set was selected and used consistently throughout the ablation experiment.

2.5.1. Tree-based importance using mean decrease of impurity (MDI)

We first used the mean decrease of impurity (MDI) to evaluate the importance of the molecular fingerprint feature, which quantifies the contribution of each feature to the reduction of impurity across all trees in the forest [36,37]. We aggregated the importance scores of each dimensionality setting from 2^3 to 2^{10} using both mean and sum strategies (Fig. 5**a**). It can be observed that lower dimensions (e.g., 32–128) of fingerprints contributed more to the predictive performance per dimension on average, while higher dimensions showed reduced contributions. This may be due to feature redundancy or sparsity in higher-dimensional representations, which dilutes the discriminative signal and thus makes it difficult to extract biologically relevant patterns from each warhead. The mean importance score reaches its peak at dimension 128, which was therefore selected as the fingerprint dimensionality for subsequent ablation analysis and final model training.

2.5.2. Tree-based importance using mean decrease of impurity (MDI)

To better understand the contribution of features to model performance, we conducted a feature ablation study, a common strategy in machine learning to assess feature importance [38,39]. As illustrated in Fig. 5**b**, we defined a sequential ablation order that includes both graph-based and conventional molecular descriptors, such as molecular fingerprints, physicochemical properties (e.g., exact mass, XLogP3), and structural features (e.g., ring count, rotatable bonds, hydrogen bond donor/acceptor counts). We first trained a baseline model using the complete set of features and then retrained models after removing each feature individually for each of the machine learning methods. The resulting performance changes reveal how each model depends on different feature types or groups (Fig. 5**c**). Removing molecular fingerprints produced divergent effects: GAT does not show too much variation when ablating this feature, so including fingerprint seems to be unnecessary for GAT. Gradient Boosting showed a small improvement, whereas logistic regression experienced a clear drop, with multilayer perceptron (MLP) and random forest also trending downward. This pattern suggests that Gradient Boosting can overfit to fingerprint features, while simpler linear models depend on them for separability. In contrast, ablating graph-specific descriptors, especially the PAGTN edge and node features, caused the largest degradations for GAT in comparison with baseline model. Notably, the removal of edge features resulted in slight performance improvements for the MLP and logistic regression models, whereas the exclusion of node features induced a performance decline in these models. This indicates that GAT relies on graph-structured representations to learn off-target patterns, whereas traditional models do not exploit these relational signals effectively.

Moreover, removing hydrogen bond donor and acceptor counts led to decreased performance for GAT model. Both features are known to associate with molecular recognition and interaction specificity [40]. Previous studies have emphasised the role of hydrogen bonding in determining ligand binding affinity and pharmacological activity [41],

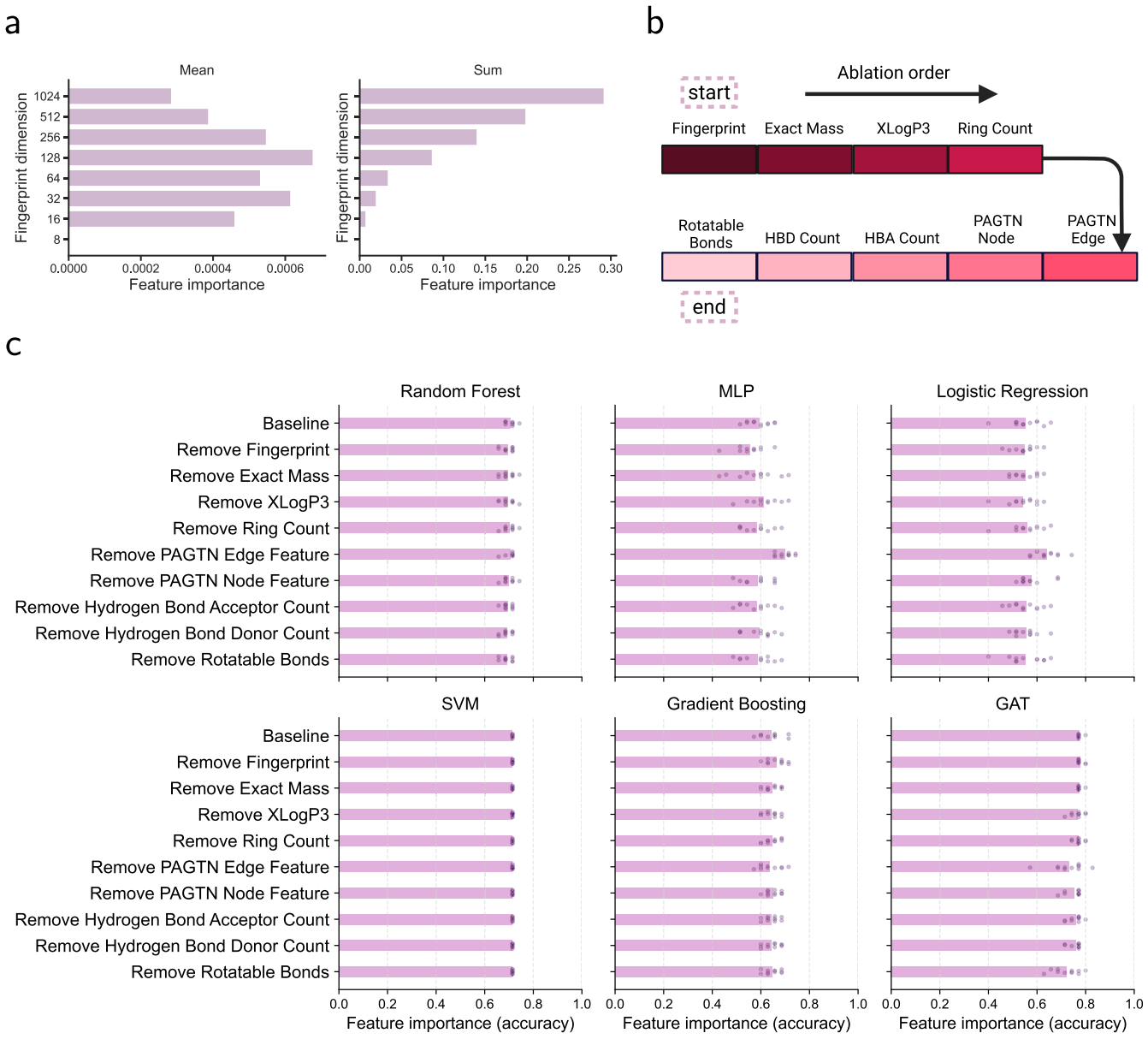

Fig. 5. Feature importance analysis for evaluating model dependence on molecular descriptors and graph-based features. (a). Mean and sum feature importance scores for molecular fingerprints of varying dimensionalities, calculated using the mean decrease in impurity (MDI) method. (b). Schematic of the predefined ablation order used in our feature removal experiments. (c). Performance impact of removing features accumulatively across various machine learning models. Bars indicate model accuracy relative to the baseline (all-feature) setup.

which may account for the assumption that models capable of capturing such interactions exhibited sensitivity to their removal. Interestingly, SVM exhibited minimal performance fluctuations across ablations, suggesting a degree of robustness or insensitivity to individual feature perturbations, reflecting the robustness and insensitivity of kernel-based learning mechanism in handling high-dimensional feature space [42,43]. In summary, this ablation study demonstrates that graph-based features are crucial for GAT, which we retained for training GAT models, while traditional models benefit more from the use of physicochemical and fingerprint-based features.

3. Methods and materials

3.1. Definition of off-target effects of PROTACs

We stratified the warhead moieties of a PROTAC molecule based on its activity levels in drug-target interactions (DTIs), which we quantified by counting the total number of unique DTIs in which the warhead appears. The underlying rationale is that warheads involved in more DTIs are likely to influence a broader range of cellular processes and disease pathways [11,44], thereby increasing the risk of off-target effects [12,45]. In this work, we defined the off-target effect of a warhead as its propensity to interact with unintended protein targets (i.e., number of its unique DTIs). Suppose a warhead is involved in N_T DTIs; it is then assigned a label y based on a group of predefined thresholds $\vec{t} = \{t_1, t_2, \ldots, t_k\}$.

$$y = \begin{cases} 0, \& N_T \le t_1 \\ 1, \& t_1 < N_T \le t_2 \\ \ldots, \& \ldots \\ k, \& t_k \le N_T \end{cases}$$

Therefore, we transformed this task into a multi-class classification problem to predict the off-target potential of a warhead based on its molecular structure and associated gene interactions, leveraging ma-

chine learning to capture both local chemical features and global interaction patterns. k was determined based on our comparative analysis.

3.2. Warhead-DTI mapping overview

To estimate warhead-induced off-target effects of PROTACs, we systematically mapped each warhead molecule from our curated warhead database to known DTIs sourced from multiple molecular interaction databases (Fig. 1**a**). This mapping was based on matching PubChem CIDs between the warhead dataset and the DTI entries (Fig. 1**b**). After establishing these links, we created an integrated dataset that associates each warhead with its corresponding interacting targets and drug source annotations. Warheads with higher DTI counts were presumed to have broader bioactivity profiles, potentially affecting more biological pathways. This provided a quantitative basis to stratify warheads into low, moderate, or high off-target potential categories to generate labels for machine learning.

3.3. PROTAC database curation

We pulled PROTAC molecules from PROTAC-DB (version 2.0: available at http://cadd.zju.edu.cn/protacdb) [24], each with its warhead and E3 ligand moieties annotated. We processed the PROTAC-DB database and deduplicated them according to the Simplified Molecular Input Line Entry System (SMILES) strings of warheads, leaving 569 unique warheads. We filtered the warheads based on the availability of valid PubChem CIDs, retaining a total of 479 compounds with identifiable chemical structures for downstream analysis.

3.4. Data reconciliation of DTIs for warhead-specific target mapping

To construct a dataset for classifying warheads based on their off-target potential, we collected DTIs from the DGIdb Database (version: December 2024, available at https://dgidb.org) [46]. We retrieved 98,923 DTIs, each containing a gene name and a composite drug identifier in the format of "source:identifier" (e.g., "ChEMBL:CHEMBL123"). The source indicates the originating database (e.g., ChEMBL [47]) and the identifier is unique within that source. This dataset serves as the foundation for linking warheads to their molecular representations. Each warhead identifier was split into a source name and a source-specific identifier. To standardise the warhead identifiers from heterogeneous sources of DGIdb and allow precise cross-referencing with warheads from PROTAC-DB, we mapped these source-specific identifiers to PubChem CIDs. This process leverages resources like PubChem annotations, external mappings, and identifier exchange services. DTIs from these data sources were retained only if PubChem CIDs were directly available in DGIdb or manually identifiable (Fig. 1**b-c**); otherwise, they were excluded from further processing, as described below.

- To map identifiers from the RxCUI database [48], we manually collected 6860 RxCUI-specific annotations from the PubChem website (https://pubchem.ncbi.nlm.nih.gov/source/25200#data=Annotations), which were originally compiled in the JSON format. Out of the 39,786 DTIs in RxCUI, we successfully discovered 12,045 DTIs with their warheads having valid PubChem CIDs.
- To map identifiers from the NCIt database [49], we first pulled 42,520 annotations from PubChem (https://pubchem.ncbi.nlm.nih.gov/source/11940#data=Annotations). To maximise data coverage, we also handled indirect mapping from NCIt to PubChem identifiers by first linking NCIt identifiers to ChEBI entries, and subsequently mapping the targeted entries to PubChem CIDs. We downloaded an NCIt-to-ChEBI mapping list from https://evs.nci.nih.gov/ftp1/NCI_Thesaurus/Mappings/NCIt-ChEBI_Mapping.txt, followed by a ChEBI-to-PubChem conversion using the Drug Identifiers Data Package available at http://go.drugbank.com/data_packages/drug_identifiers. The two-step mapping resulted in 1241 DTIs with valid PubChem CIDs.
- To map identifiers from the IUPHAR.ligand database [50], we sourced a complete ligand list from the IUPHAR/BPS Guide to PHARMACOLOGY (https://www.guidetopharmacology.org/download.jsp#data). Out of 15,085 DTIs, 12,853 warheads with explicit PubChem CID annotations in the Guide were directly adopted. For unmapped identifiers, no further mapping was pursued due to a rather limited number of annotation resources.
- To map identifiers from the ChEMBL database [47], we employed the PubChem Identifier Exchange Service (https://pubchem.ncbi.nlm.nih.gov/idexchange/idexchange.cgi) to query PubChem CIDs of a submitted list of ChEMBL IDs. Out of 13,218 DTIs, we obtained 11,847 DTIs with valid PubChem CIDs of their warheads.
- To map identifiers from the DrugBank database [51], we extracted 16,582 precomputed mappings between DrugBank identifiers and PubChem CIDs from the Drug Identifiers Data Package (http://go.drugbank.com/data_packages/drug_identifiers). Out of 3866 DTIs, 2614 were left after successful matches.
- To map identifiers from the Wikidata database [52], we utilised the PubChem Identifier Exchange Service (https://pubchem.ncbi.nlm.nih.gov/idexchange/idexchange.cgi) to retrieve the corresponding PubChem CIDs. Out of the 117 DTIs queried, 106 were successfully matched to valid PubChem CIDs.
- For 47 DTIs from other sources (e.g., HemOnc, Drugs@FDA.NDA), no systematic mapping was feasible due to limited annotation coverage or mapping exchange services. However, we manually mapped two out of three DTIs from the ChemIDplus source, with the warheads matching known PubChem identifiers.

3.5. Intersection of PROTAC warheads with DTI data

To establish a functional link between PROTAC-derived warheads and those involved in known DTIs, we performed an exact matching of PubChem CIDs across the two sets. Specifically, we intersected the warheads curated from PROTAC-DB with those found in the DTI databases by matching their PubChem compound identifiers. Only DTIs with warheads present in both sources were retained. This filtering step yielded a final set of 172 warheads with both confirmed PROTAC relevance and associated DTI annotations, enabling further analysis of their target interaction profiles.

3.6. Quantification of warhead-induced off-target potential based on DTI diversity

For each of the 172 PROTAC warheads, we computed the number of associated targets based on their involvement in the retained DTIs above. To ensure biological interpretability, we rigorously deduplicated target entries within the interaction list of each warhead, counting only unique targets. This target count serves as a proxy for the warhead-induced off-target potential of PROTACs, with higher values indicating broader and less specific binding behavior. The resulting distribution of target counts per warhead ranged from 1 to 115.

3.7. Classification of off-target potential

We conducted a benchmarking analysis to determine the optimal thresholding strategy for assigning discrete labels that reflect the propensity of warheads to induce off-target effects. By evaluating various threshold schemes for converting continuous target counts into categorical labels, we assessed their impact on the performance of downstream multi-class classification models. Details are shown in section *Results*.

3.8. Dataset generation

Among the 172 warheads, we first randomly selected 35 (approximately 20 %) as a held-out test set according to a 1:4 ratio. The remaining 137 warheads were used for training and cross-validation. During five-fold cross-validation, these 137 samples were split into 109 for training and 28 for validation in each fold.

To systematically evaluate the generalizability of machine learning models in small molecule prediction, we employed four different clustering-based dataset splitting strategies (Random, Butina [53], UMAP [54], and Scaffold), followed by group 5-fold cross-validation (80 % training, 20 % testing in each fold). This approach aims to avoid overfitting and overly optimistic evaluations caused by high similarity between training and test molecules, a common concern in drug discovery model benchmarking.

First, it is observed that UMAP clustering results in considerable variation in the number of test molecules across different k-fold splits in its effort to maximise dissimilarity between training and test molecules (Supplementary Figure 2). The Butina and Scaffold clustering methods cause fluctuations in the number of test molecules across different folds. Further t-SNE analysis confirmed that UMAP clustering introduced a more fragmented distribution of test small molecules compared to the other strategies, which provided more uniform sampling (Supplementary Figure 3). These findings suggest that although UMAP splitting maximises dissimilarity, it may compromise training sufficiency and is therefore less suitable for robust model evaluation.

We then calculated the Tanimoto similarity [55] between small molecules in the training and test datasets under each splitting strategy (Supplementary Figure 4). While UMAP-based clustering achieved the lowest average similarity, it did so by disproportionately sampling test molecules from sparse regions of chemical space. This may lead to low coverage of specific chemical subspaces during training. In contrast, the random, Butina, and Scaffold clustering methods achieve a more balanced trade-off between dataset sampling diversity and molecular similarity. Among the three methods, Butina clustering performed best at reducing similarity between training and test molecules, because it groups molecules by Tanimoto similarity per se and keeps those below a set threshold (0.65) in separate clusters. But this method can also intensively sample data from a certain chemical space (Supplementary Figure 2). The random and Scaffold clustering methods maintain a Tanimoto similarity of below 0.6 between almost all test molecules and those required for training, with the majority sharing ~0.3 similarity. Established studies suggest that a pair of molecules is declared dissimilar if they share a Tanimoto score below 0.6 [34,56–59]. Given the limitation in our data size and the relatively low similarity between molecules in our dataset, we finally leverage the Random clustering-based dataset splits to determine model parameters. This ensured that the evaluation avoided information leakage from highly similar compounds, while still allowing the model to be fairly trained across representative chemical regions.

3.9. Graph representation of warheads

We represent each warhead as a graph $G = (V, E)$, where V denotes the set of n nodes (atoms) and E denotes the set of edges (bonds). Each node $v \in V$ is characterised by a feature vector $\boldsymbol{x} \in \mathbb{R}^{d_v}$ that encodes atom-level properties, such as its atomic number, formal charge, and chirality. Each edge $(v, u) \in E$ is represented by a feature vector $e_{u,v} \in \mathbb{R}^{d_e}$ that encodes bond-level properties, such as its bond type, conjugation, and stereochemical information. The neighborhood of node v is denoted as $N(v) = \{u | (v, u) \in E\}$. To generate these graph representations for each molecular structure, we employed path-augmented graph transformer networks (PAGTNs) [32], which transforms SMILES strings into directed graphs, including a feature matrix $\boldsymbol{X} \in \mathbb{R}^{|V| \times d_v}$ across all nodes and a feature matrix $\boldsymbol{E} \in \mathbb{R}^{|E| \times d_e}$ across all edges. This representation retains both the local atomic environments and the global molecular topology, offering a chemically meaningful foundation for downstream processing by graph neural networks. The incorporation of edge features is especially important, as it captures not only direct bonding relationships but also long-range atomic interactions through path-based augmentation.

In addition, we generated a handful of non-graph-based features using RDKit (version: 2024.9.6) and encoded them into a feature vector $f_h \in \mathbb{R}^{d_m}$. This setting can also help to determine how much graph-based and non-graph-based features contribute to the final predictive performance, respectively.

3.10. The architecture of graph neural networks

We employed a graph attention network (GAT) [60] to learn off-target effects of PROTACs from the above features. As can be seen in Fig. 2**a**, at the core of this GAT are sequentially connected two GAT convolutional layers, each followed by the ReLU activation function for fast training by introducing non-linearity while avoiding vanishing gradients [61].

Let $\boldsymbol{h}^{(l,s)} = \left\{ h_v^{(l,s)} | l \in \mathbb{N}^2, s \in \mathbb{N}^S, v \in \mathbb{N}^n \right\}$ be a set of transformed node features for warhead s at layer l. $h_v^{(l,s)} \in \mathbb{R}^{d_v}$ where d_v represents the dimension of the feature vector for node v. When $l = 0$, $\boldsymbol{h}^{(l,s)} = \boldsymbol{x}^s$. For any two nodes v and u, we denote an overall weight matrix $W = [W_v^{(l,s)} || W_u^{(l,s)}]$ at a GAT convolutional layer according to [62], where $W_v^{(l,s)} \in \mathbb{R}^{d_v \times d_h^{(l)}}$ and $W_u^{(l,s)} \in \mathbb{R}^{d_u \times d_h^{(l)}}$. $d_h^{(l)}$ is the number of neurons at layer l. Since edge $e_{u,v}$ between nodes v and u is characterised by a feature vector, we estimated an additional weight matrix $W_e^{(l,s)} \in \mathbb{R}^{d_e \times d_h^{(l)}}$ for these edge features. Then, the overall weight matrix W is rewritten as the concatenation of the three submatrices, $[W_v^{(l,s)} || W_u^{(l,s)} || W_e^{(l,s)}]$. The attention coefficient $c_{v,u}^{(l,s)}$ at layer l between nodes v and u is computed by

$$c_{v,u}^{(l,s)} = \frac{\exp(e\left(h_v^{(l,s)}, h_u^{(l,s)}\right))}{\sum_{r \in N(v)} e\left(h_v^{(l,s)}, h_r^{(l,s)}\right)}$$

where e represents a function that aggregates the features of node v and those of any neighboring node u, which are parameterised by two weight matrices $W_v^{(l,s)}$ and $W_u^{(l,s)}$, respectively. For node v, the attention mechanism is made at layer l by setting a group of neurons $a_v^{(l,s)} \in \mathbb{R}^{d_h^{(l)}}$ to parameterise the transformed input (Fig. 2**b**). e is defined in the following form

$$e\left(h_v^{(l,s)}, h_u^{(l,s)}\right) = \psi(a_v^{(l,s)T} W_v^{(l,s)} h_v^{(l,s)} + a_u^{(l,s)T} W_u^{(l,s)} h_u^{(l,s)} + a_e^{(l,s)T} W_e^{(l,s)} e_u^{(l,s)})$$

where ψ is the Leaky ReLU activation function to prevent neuron inactivity by setting negative input to non-zero slope [63].

The $h_v^{(l,s)}$ at layer l is updated to $h_v^{(l+1,s)}$ at layer $l+1$ using the following equation.

$$h_v^{(l+1,s)} = \sum_{u \in N(v)} c_{v,u}^{(l,s)} W_u^{(l,s)} h_u^{(l,s)}$$

We then performed a global mean pooling operation on the output feature vectors $\boldsymbol{h}^{(l,s)}$ to potentially help capture the most significant signals across the graph.

$$z = GlobalMeanPool(\boldsymbol{h}^{(l,s)})$$

Then, z is further parameterised using a fully-connected neural network (FNN) composed of a single layer with 64 neurons, denoted as FNN_1.

$$z' = FNN_1(z) \in \mathbb{R}^{64}$$

To obtain embeddings of non-graph-based features f_h, we used a separate FNN (denoted as FNN_2) that consists of two hidden layers each configured with 64 neurons (Fig. 2**c**).

$$f_h^{'} = FNN_2(f_h) \in \mathbb{R}^{64}$$

The rear portion of our GAT-based deep learning architecture is an FNN composed of a single layer with 64 neurons, followed by a multi-class softmax function for converting logits into a vector of probabilities $\boldsymbol{p}$, such that

$$\boldsymbol{p} = softmax(FNN_3(z^{'}||f_h^{'}))$$

For comparison, we also implemented a graph convolutional network (GCN) as a baseline model, using the same two-layer architecture as GAT but without the attention mechanism.

3.11. Model comparison

To evaluate the impact of the attention mechanism on prediction performance, we compared graph convolutional neural network models with and without attention against a set of commonly used machine learning algorithms reported in the literature, including random forest [36], gradient boosting [64], support vector machine (SVM) [65], logistic regression [66], and multilayer perceptron (MLP) [67]. These baseline models were implemented using the Scikit-learn library [68]. These classical baselines (random forest (RF), gradient boosting (GB), SVM, logistic regression (LR), MLP) use the same underlying graph representation as the GNNs. For each warhead, PAGTN produces 94-dimensional node features and 42-dimensional edge features. We summarise these matrices by computing per-dimension mean, standard deviation, skewness, and kurtosis over nodes and over edges. Then we concatenate the results to form a 544-dimensional moment-pooled graph vector ($4 \times 94 + 4 \times 42$). This fixed-length vector is standardised within each training fold and fed to the classical models. Unless explicitly stated as an ablation, handcrafted physicochemical descriptors are not used. In ablation experiments only, we concatenate nine descriptors (listed in Supplementary Table S5-S11) to the moment-pooled vector to quantify their incremental contribution. All baselines share the same train/validation/test splits and preprocessing as GAT/GCN. Results for baselines are reported in Supplementary Tables S5–S11.

3.12. Model setup

The GAT models were trained using the negative log-likelihood loss with the Adam optimiser for 100 epochs. The learning rate was set to 0.001 and the dropout rate was set to 0.5. Warheads were fed into the models with a batch size of 32. Model training experiments were conducted using PyTorch [69]. Other machine learning models were trained using Scikit-learn with parameters by default [68]. The random forest model was trained and its hyperparameters were fine-tuned via grid search. After fine-tuning, the number of trees is set to 100 and the maximum depth is set to 10. All models were tested to determine the number of labels (i.e., $\vec{t}$) for multi-class classification. Five-fold cross-validation (cv1–cv5) was performed on the training set, with results averaged to compute mean and standard deviation where applicable. The test set (35 warheads) was held out for final evaluation.

3.13. Evaluation metrics

Model performance was assessed using threshold-based metrics, including accuracy, precision, and recall, F1-score, specificity, and Jaccard index, which were computed as weighted averages across multi-classes to account for potential class imbalance. In addition, we adopted two threshold-free metrics, AUCPR, which measures the area under the precision-recall curve and averages precision across recall levels, and AUC, which measures the area under the receiver operating characteristic (ROC) curve. For cross-validation, we report mean performance across folds.

3.14. SENTINEL

According to our evaluation based on feature ablation experiments, we finally only retained graph-based features as input to the GAT architecture to train models. The model, with parameters determined strictly via cross-validation, was termed SENTINEL and retained for public use.

4. Discussion

In this study, we introduced SENTINEL, a deep learning framework to assess warhead-induced off-target potential of PROTACs based on their involvement levels in known drug–target interactions. Our study aims to reinforce the value of early predictive assessment of off-target effects as a critical component of rational PROTAC design. As the PROTAC modality continues to expand into new therapeutic areas, minimising unintended protein degradation is paramount to ensuring safety and specificity. By using the number of unique protein targets associated with each warhead as a surrogate marker for off-target potential, we provide an interpretable method for early-stage risk assessment of PROTAC components. The adoption of PAGTN-encoded graph representations of small molecules, coupled with GATs for predictive learning, allowed the model to effectively capture both local and global structural features relevant to warhead behaviour. Notably, while GAT and GCN models achieved competitive performance on the held-out test set, their performance during cross-validation was relatively low. This discrepancy likely stems from the limited size of the available training data. Future directions include increasing data size, expanding the dataset with experimental degradation profiles, refining risk metrics with functional annotations, and applying the framework to broader classes of heterobifunctional degraders beyond PROTACs.

Ethics information

The work analyzes publicly available, de-identified datasets. Use complied with the respective data-use licenses. No human and no animal experiments were performed.

CRediT authorship contribution statement

Yutong Hu: Writing – original draft, Visualization, Investigation, Formal analysis, Data curation. **Kieran Didi:** Methodology, Investigation. **Adam P. Cribbs:** Resources, Funding acquisition. **Jianfeng Sun:** Writing – review & editing, Writing – original draft, Supervision, Methodology, Investigation, Funding acquisition, Data curation, Conceptualization.

Declaration of Competing Interest

Jianfeng Sun is the founder and operator of MinAI, a commercially run platform for artificial intelligence-based discovery of drugs targeting proteins and non-coding RNAs. Adam Cribbs is listed as an inventor on several patents filed by Oxford University Innovations concerning single-cell sequencing technologies. Other authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgement

We are grateful to Professor Alex Bullock for constructive discussions. A.P.C. and J.S. were supported by the Medical Research Council (MRC) career development fellowship (MR/V010182/1).

Code and data avail ability

The program is publicly available at https://github.com/mivolis/SENTINEL.

Appendix A. Supporting information

Supplementary data associated with this article can be found in the online version at doi:10.1016/j.csbj.2025.10.028.

References

[1] Pettersson M, Crews CM. PROteolysis TArgeting chimeras (PROTACs) — past, present and future. Drug Discov Today Technol 2019;31:15–27. https://doi.org/10.1016/j.ddtec.2019.01.002.

[2] Adjei AA. What is the right dose? The elusive optimal biologic dose in phase I clinical trials. J Clin Oncol 2006;24:4054–5. https://doi.org/10.1200/JCO.2006.07.4658.

[3] Zhao L, Zhao J, Zhong K, Tong A, Jia D. Targeted protein degradation: mechanisms, strategies and application. Signal Transduct Target Ther 2022;7:113. https://doi.org/10.1038/s41392-022-00966-4.

[4] Sun X, Gao H, Yang Y, He M, Wu Y, Song Y, et al. PROTACs: great opportunities for academia and industry. Signal Transduct Target Ther 2019;4:64. https://doi.org/10.1038/s41392-019-0101-6.

[5] Huang X, Dixit VM. Drugging the undruggables: exploring the ubiquitin system for drug development. Cell Res 2016;26:484–98. https://doi.org/10.1038/cr.2016.31.

[6] Nalawansha DA, Crews CM. PROTACs: an emerging therapeutic modality in precision Medicine. Cell Chem Biol 2020;27:998–1014. https://doi.org/10.1016/j.chembiol.2020.07.020.

[7] Hsia O, Hinterndorfer M, Cowan AD, Iso K, Ishida T, Sundaramoorthy R, et al. Targeted protein degradation via intramolecular bivalent glues. Nature 2024;627:204–11. https://doi.org/10.1038/s41586-024-07089-6.

[8] Li X, Pu W, Zheng Q, Ai M, Chen S, Peng Y. Proteolysis-targeting chimeras (PROTACs) in cancer therapy. Mol Cancer 2022;21:99. https://doi.org/10.1186/s12943-021-01434-3.

[9] Liu J, Ma J, Liu Y, Xia J, Li Y, Wang ZP, et al. PROTACs: a novel strategy for cancer therapy. Semin Cancer Biol 2020;67:171–9. https://doi.org/10.1016/j.semcancer.2020.02.006.

[10] Schapira M, Calabrese MF, Bullock AN, Crews CM. Targeted protein degradation: expanding the toolbox. Nat Rev Drug Discov 2019;18:949–63. https://doi.org/10.1038/s41573-019-0047-y.

[11] Cromm PM, Crews CM. Targeted protein degradation: from chemical biology to drug discovery. Cell Chem Biol 2017;24:1181–90. https://doi.org/10.1016/j.chembiol.2017.05.024.

[12] Bondeson DP, Smith BE, Burslem GM, Buhimschi AD, Hines J, Jaime-Figueroa S, et al. Lessons in PROTAC design from selective degradation with a promiscuous warhead. Cell Chem Biol 2018;25:78–87.e5. https://doi.org/10.1016/j.chembiol.2017.09.010.

[13] Quach H, Ritchie D, Stewart AK, Neeson P, Harrison S, Smyth MJ, et al. Mechanism of action of immunomodulatory drugs (IMiDS) in multiple myeloma. Leukemia 2010;24:22–32. https://doi.org/10.1038/leu.2009.236.

[14] Nguyen TM, Sreekanth V, Deb A, Kokkonda P, Tiwari PK, Donovan KA, et al. Proteolysis-targeting chimeras with reduced off-targets. Nat Chem 2024;16:218–28. https://doi.org/10.1038/s41557-023-01379-8.

[15] Shen C, Nayak A, Neitzel LR, Adams AA, Silver-Isenstadt M, Sawyer LM, et al. The E3 ubiquitin ligase component, cereblon, is an evolutionarily conserved regulator of wnt signaling. Nat Commun 2021;12:5263. https://doi.org/10.1038/s41467-021-25634-z.

[16] Guo H. Bumped pomalidomide-based PROTACs. Commun Chem 2024;7:41. https://doi.org/10.1038/s42004-024-01125-2.

[17] Wu W, Nelson GM, Koch R, Donovan KA, Nowak RP, Heavican-Foral TB, et al. Overcoming IMiD resistance in T-cell lymphomas through potent degradation of ZFP91 and IKZF1. Blood 2022;139:2024–37. https://doi.org/10.1182/blood.2021014701.

[18] Winter GE, Buckley DL, Paulk J, Roberts JM, Souza A, Dhe-Paganon S, et al. Phthalimide conjugation as a strategy for in vivo target protein degradation. Science 2015;348(1979):1376–81. https://doi.org/10.1126/science.aab1433.

[19] Wang C, Zhang Y, Yang S, Chen W, Xing D. PROTACs for BRDs proteins in cancer therapy: a review. J Enzym Inhib Med Chem 2022;37:1694–703. https://doi.org/10.1080/14756366.2022.2081164.

[20] Bond AG, Craigon C, Chan K-H, Testa A, Karapetsas A, Fasimoye R, et al. Development of BromoTag: a "Bump-and-Hole"–PROTAC system to induce potent, rapid, and selective degradation of tagged target proteins. J Med Chem 2021;64:15477–502. https://doi.org/10.1021/acs.jmedchem.1c01532.

[21] Bondeson DP, Mares A, Smith IED, Ko E, Campos S, Miah AH, et al. Catalytic in vivo protein knockdown by small-molecule PROTACs. Nat Chem Biol 2015;11:611–7. https://doi.org/10.1038/nchembio.1858.

[22] Burslem GM, Crews CM. Proteolysis-targeting chimeras as therapeutics and tools for biological discovery. Cell 2020;181:102–14. https://doi.org/10.1016/j.cell.2019.11.031.

[23] Peteani G, Huynh MTD, Gerebtzoff G, Rodríguez-Pérez R. Application of machine learning models for property prediction to targeted protein degraders. Nat Commun 2024;15:5764. https://doi.org/10.1038/s41467-024-49979-3.

[24] Ge J, Li S, Weng G, Wang H, Fang M, Sun H, et al. PROTAC-DB 3.0: an updated database of PROTACs with extended pharmacokinetic parameters. Nucleic Acids Res 2025;53:D1510–5. https://doi.org/10.1093/nar/gkae768.

[25] Li F, Hu Q, Zhang X, Sun R, Liu Z, Wu S, et al. DeepPROTACs is a deep learning-based targeted degradation predictor for PROTACs. Nat Commun 2022;13:7133. https://doi.org/10.1038/s41467-022-34807-3.

[26] Li F, Hu Q, Zhou Y, Yang H, Bai F. DiffPROTACs is a deep learning-based generator for proteolysis targeting chimeras. Brief Bioinform 2024;25. https://doi.org/10.1093/bib/bbae358.

[27] Abbas A, Ye F. Computational methods and key considerations for in silico design of proteolysis targeting chimera (PROTACs). Int J Biol Macromol 2024;277:134293. https://doi.org/10.1016/j.ijbiomac.2024.134293.

[28] Zhang W, Roy Burman SS, Chen J, Donovan KA, Cao Y, Shu C, et al. Machine learning modeling of Protein-Intrinsic features predicts tractability of targeted protein degradation. Genom Proteom Bioinforma 2022;20:882–98. https://doi.org/10.1016/j.gpb.2022.11.008.

[29] Cai L, Yue G, Chen Y, Wang L, Yao X, Zou Q, et al. ET-PROTACs: modeling ternary complex interactions using cross-modal learning and ternary attention for accurate PROTAC-induced degradation prediction. Brief Bioinform 2024;26. https://doi.org/10.1093/bib/bbae654.

[30] Ribes S, Nittinger E, Tyrchan C, Mercado R. Modeling PROTAC degradation activity with machine learning. Artif Intell Life Sci 2024;6:100104. https://doi.org/10.1016/j.ailsci.2024.100104.

[31] Wang L., Jinsong S., Gong Q., Yin Z., Chen Y., Hao Y., et al. LM-PROTAC: a language model-driven PROTAC generation pipeline with dual constraints of structure and property 2025. https://doi.org/10.21203/rs.3.rs-6356959/v1.

[32] Chen B., Barzilay R., Jaakkola T. Path-augmented graph transformer network. ArXiv Preprint ArXiv:190512712 2019.

[33] Yun S, Jeong M, Kim R, Kang J, Kim HJ. Graph transformer networks. In: Wallach H, Larochelle H, Beygelzimer A, d Alché-Buc F, Fox E, Garnett R, editors. Adv Neural Inf Process Syst, 32. Curran Associates, Inc; 2019.

[34] You J, McLeod RD, Hu P. Predicting drug-target interaction network using deep learning model. Comput Biol Chem 2019;80:90–101. https://doi.org/10.1016/j.compbiolchem.2019.03.016.

[35] Greener JG, Kandathil SM, Moffat L, Jones DT. A guide to machine learning for biologists. Nat Rev Mol Cell Biol 2022;23:40–55. https://doi.org/10.1038/s41580-021-00407-0.

[36] Breiman L. Random forests. Mach Learn 2001;45:5–32. https://doi.org/10.1023/A:1010933404324.

[37] Louppe G. Understanding Random Forests: From Theory to Practice. ArXiv 2015.

[38] Zeng M, Li M, Fei Z, Wu F-X, Li Y, Pan Y, et al. A deep learning framework for identifying essential proteins by integrating multiple types of biological information. IEEE/ACM Trans Comput Biol Bioinform 2021;18:296–305. https://doi.org/10.1109/TCBB.2019.2897679.

[39] Yang Q, Sun J, Wang X, Wang J, Liu Q, Ru J, et al. SVLearn: a dual-reference machine learning approach enables accurate cross-species genotyping of structural variants. Nat Commun 2025;16:2406. https://doi.org/10.1038/s41467-025-57756-z.

[40] Patel K. Structural organization of peptides. De Novo Peptide Design. Elsevier; 2023. p. 1–33. https://doi.org/10.1016/B978-0-323-99917-5.00008-1.

[41] Bauer CA, Schneider G, Göller AH. Machine learning models for hydrogen bond donor and acceptor strengths using large and diverse training data generated by first-principles interaction free energies. J Chemin 2019;11:59. https://doi.org/10.1186/s13321-019-0381-4.

[42] Torkamani M.A., Lowd D. On Robustness and Regularization of Structural Support Vector Machines. In: Xing EP, Jebara T, editors. Proceedings of the 31st International Conference on Machine Learning, vol. 32, Bejing, China: PMLR; 2014, p. 577–585.

[43] Quadrianto N, Lampert CH. Learning, Kernel-based. Encyclopedia of Systems Biology. New York, NY: Springer New York; 2013. p. 1114–7. https://doi.org/10.1007/978-1-4419-9863-7_604.

[44] Liu Z, Hu M, Yang Y, Du C, Zhou H, Liu C, et al. An overview of PROTACs: a promising drug discovery paradigm. Mol Biomed 2022;3:46. https://doi.org/10.1186/s43556-022-00112-0.

[45] Ábrányi-Balogh P, Petri L, Imre T, Szijj P, Scarpino A, Hrast M, et al. A road map for prioritizing warheads for cysteine targeting covalent inhibitors. Eur J Med Chem 2018;160:94–107. https://doi.org/10.1016/j.ejmech.2018.10.010.

[46] Cannon M, Stevenson J, Stahl K, Basu R, Coffman A, Kiwala S, et al. DGIdb 5.0: rebuilding the drug–gene interaction database for precision Medicine and drug discovery platforms. Nucleic Acids Res 2024;52:D1227–35. https://doi.org/10.1093/nar/gkad1040.

[47] Mendez D, Gaulton A, Bento AP, Chambers J, De Veij M, Félix E, et al. ChEMBL: towards direct deposition of bioassay data. Nucleic Acids Res 2019;47:D930–40. https://doi.org/10.1093/nar/gky1075.

[48] Nelson SJ, Zeng K, Kilbourne J, Powell T, Moore R. Normalized names for clinical drugs: RxNorm at 6 years. J Am Med Inform Assoc 2011;18:441–8. https://doi.org/10.1136/amiajnl-2011-000116.

[49] de Coronado S., Remennik L., Elkin P.L. National Cancer Institute Thesaurus (NCIt), 2023, p. 395–441. https://doi.org/10.1007/978-3-031-11039-9_17.

[50] Harding SD, Armstrong JF, Faccenda E, Southan C, Alexander SPH, Davenport AP, et al. The IUPHAR/BPS guide to PHARMACOLOGY in 2024. Nucleic Acids Res 2024;52:D1438–49. https://doi.org/10.1093/nar/gkad944.

[51] Knox C, Wilson M, Klinger CM, Franklin M, Oler E, Wilson A, et al. DrugBank 6.0: the DrugBank knowledgebase for 2024. Nucleic Acids Res 2024;52:D1265–75. https://doi.org/10.1093/nar/gkad976.
[52] Vrandečić D, Krötzsch M. Wikidata: a free collaborative knowledgebase. Commun ACM 2014;57:78–85. https://doi.org/10.1145/2629489.
[53] Butina D. Unsupervised data base clustering based on daylight's fingerprint and tanimoto similarity: a fast and automated way to cluster small and large data sets. J Chem Inf Comput Sci 1999;39:747–50. https://doi.org/10.1021/ci9803381.
[54] Guo Q, Hernandez-Hernandez S, Ballester PJ. UMAP-based clustering split for rigorous evaluation of AI models for virtual screening on cancer cell lines*. J Chemin 2025;17:94. https://doi.org/10.1186/s13321-025-01039-8.
[55] Bajusz D, Rácz A, Héberger K. Why is tanimoto index an appropriate choice for fingerprint-based similarity calculations? J Chemin 2015;7:20. https://doi.org/10.1186/s13321-015-0069-3.
[56] Huber F, van der Burg S, van der Hooft JJJ, Ridder L. MS2DeepScore: a novel deep learning similarity measure to compare tandem mass spectra. J Chemin 2021;13:84. https://doi.org/10.1186/s13321-021-00558-4.
[57] Kuwahara H, Gao X. Analysis of the effects of related fingerprints on molecular similarity using an eigenvalue entropy approach. J Chemin 2021;13:27. https://doi.org/10.1186/s13321-021-00506-2.
[58] Bui-Thi D, Liu Y, Lippens JL, Laukens K, De Vijlder T. TransExION: a transformer based explainable similarity metric for comparing IONS in tandem mass spectrometry. J Chemin 2024;16:61. https://doi.org/10.1186/s13321-024-00858-5.
[59] Strobel M, Gil-de-la-Fuente A, Zare Shahneh MR, Abiead Y, El, Bushuiev R, Bushuiev A, et al. An evaluation methodology for machine learning-based tandem mass spectra similarity prediction. BMC Bioinforma 2025;26:174. https://doi.org/10.1186/s12859-025-06194-1.
[60] Veličković P, Cucurull G, Casanova A, Romero A, Liò P, Bengio Y. Graph attention networks. Int Conf Learn Represent 2018.
[61] Nair V., Hinton G.E. Rectified Linear Units Improve Restricted Boltzmann Machines. Proceedings of the 27th International Conference on International Conference on Machine Learning, Madison, WI, USA: Omnipress; 2010, p. 807–814.
[62] Brody S, Alon U, Yahav E. How attentive are graph attention networks? Int Conf Learn Represent 2022.
[63] Xu B., Wang N., Chen T., Li M. Empirical Evaluation of Rectified Activations in Convolutional Network. ArXiv 2015.
[64] Miller A, Panneerselvam J, Liu L. A review of regression and classification techniques for analysis of common and rare variants and gene-environmental factors. Neurocomputing 2022;489:466–85. https://doi.org/10.1016/j.neucom.2021.08.150.
[65] Cortes C, Vapnik V. Support-vector networks. Mach Learn 1995;20:273–97. https://doi.org/10.1007/BF00994018.
[66] Pathan S, Prabhu KG, Siddalingaswamy PC. Techniques and algorithms for computer aided diagnosis of pigmented skin lesions—a review. Biomed Signal Process Control 2018;39:237–62. https://doi.org/10.1016/j.bspc.2017.07.010.
[67] Gardner MW, Dorling SR. Artificial neural networks (the multilayer perceptron)—a review of applications in the atmospheric sciences. Atmos Environ 1998;32:2627–36. https://doi.org/10.1016/S1352-2310(97)00447-0.
[68] Pedregosa F, Varoquaux G, Gramfort A, Michel V, Thirion B, Grisel O, Blondel M, et al. Scikit-learn: machine learning in python. J Mach Learn Res 2011;12:2825–30.
[69] Paszke A., Gross S., Massa F., Lerer A., Bradbury J., Chanan G., et al. PyTorch: an imperative style, high-performance deep learning library. Proceedings of the 33rd International Conference on Neural Information Processing Systems, Red Hook, NY, USA: Curran Associates Inc.; 2019.