Electronic Research Archive

http://www.aimspress.com/journal/era

ERA, 34(1): 69–89.
DOI: 10.3934/era.2026005
Received: 20 September 2025
Revised: 12 November 2025
Accepted: 24 November 2025
Published: 29 December 2025

*Research article*

# Gated graph neural network-based features for epigenetic target detection

**Fengwei Jing[1], Yuxuan Liu[2], Xinhui Si[2] and Yishu Wang[2,*]**

[1] National Engineering Research Center for Advanced Rolling and Intelligent Manufacturing, University of Science and Technology Beijing, Beijing 100083, China

[2] School of Mathematics and Statistics, University of Science and Technology Beijing, Beijing 100083, China

* **Correspondence:** Email: yishu6661@126.com.

**Abstract:** Epigenetics refers to genetic modifications that cause changes in gene expression and are not encoded in DNA. These modifications, including histone, DNA, and RNA, as well as miRNA, are closely related to oncology research and have the potential to influence the treatment of metabolic, neurological, inflammatory, and cardiovascular diseases. In cancer treatment, targeting abnormal epigenetic regulation can be achieved through inhibition of abnormal DNA methylation and histone methylation and acetylation. However, the discovery of efficient epigenetic drugs often requires substantial time and resources. To aid in drug research, we propose an extraction method of molecule features based on a gated graph neural network (GGNN) and integrate four supervised classifiers to predict epigenetic targets in screened chemogenomic datasets. Our GGNN+XGBoost integrated model achieves an overall accuracy of 0.8158, with accuracy increasing to 0.981 in certain target datasets. This approach offers a promising solution for predicting epigenetic targets, with implications for cancer treatment and other diseases. Furthermore, our method showcases the potential of utilizing big datasets and predictive models to expedite epigenetic drug discovery.

**Keywords:** GGNN; XGBoost; epigenetic targets; supervised classifiers; drug discovery

## 1. Introduction

The epigenotype was first named by Dr. Waddington [1] to define the phenotypic changes occurring without genotype alteration and to explain various aspects of its development. About 75 years, we discovered that epigenetic control of gene expression patterns is transmitted not by changes in the DNA sequence but in the state of chromatin, the physiological form of genetic information. In addition to DNA, epigenetic mechanisms also stabilize gene expression programs to determine cell types.

Today, epigenetics is an extensive field of study involving morphogenesis, cell heredity, transgenerational epigenetic inheritance, and evo-devo approaches. The interactions between proteins, changes in protein phosphorylation, histone modification, and DNA methylation are all cell responses to the developmental environment. Esteller and others demonstrated that these adaptive changes producing permanent alterations might lead to pathology later in life [2–4]. Thus, epigenetic dysregulation has been associated with altered gene expression patterns and multiple diseases [5]. Accordingly, the prediction of epigenetic targets is an important process in drug discovery and an emerging strategy for chronic and complex diseases. The increased interest in epigenetics requires the development and maintenance of large datasets of structure-epigenetic activity relationships for epigenetic targets.

Cancer research typically focuses on genetic variations such as mutations, gene rearrangements, and copy number variations that occur during tumorigenesis. Early on, the discovery of abnormal epigenetic features (such as DNA methylation) of potential clinical importance in cancer provided epigenetic treatments such as chemical inhibitors of DNA methylation (DNMTi), (DNMT: DNA methyltransferase) [6]; readers which reactivate abnormally silenced tumor suppressant genes, combined with HDACi (HDAC: histone deacetylase) [7]; erasers such as treostatin A (TSA) and Trapoxin.

In 2006, the first epigenetic drugs (decitabine and vorinota) for the treatment of human cancers were approved by the US Food and Drug Administration (FDA), which provide one compelling argument for treating cancer by reversing epigenetic errors. Histone-modifying enzymes, both reader and eraser, are potential drug targets in oncology. Filippakopoulos et al. [8] added new readers such as the bromodomain and extra-terminal (BET) family similar with hypoxantin (HAT) and HDAC inhibitors, one of which targets BRD4 (BRD: bromodomain), a protein critical to the expression of tumor promoting key oncogenes (e.g., myelocytomatosis viral oncogene homolog (MYC)) and pro-inflammatory genes (e.g., nuclear factor kappa-B (NFKB)) in blood cancer. Allis and Jenuwein [9] also demonstrated that epigenetic control is important for immune cell activation and can enhance immune responses; for example, HDAC inhibitors can maintain T cell activation by preventing activation-induced cell death [10].

However, as indicated by Glozak et al. [11], nearly histone-modifying enzymes target many non-histone nuclear and cytoplasmic proteins; hence, the challenge of cancer therapy through epigenetic drugs is to carefully predict and use epigenetic targets. Efficient epigenetic drug discovery and development requires high-quality tools, including recombinant proteins, enzymes, effective antibodies, and other experimental approaches. In addition, during screening, the correct protein conformation is critical; for example, when screening and identifying small molecules that inhibit epigenetic related enzymes, only the catalytic domain is required, while for screening advanced structural complex protein states, recombinant proteins are likely required. Despite the wealth of important chemogenomic databases [12–15], such as EpiDBase, an interactive database of small molecule ligands of epigenetic protein families including 5784 unique ligands of various epigenetic markers [15], these large rich structure-activity relationships datasets lack effective predictive models. Only a previous descriptive study explored structure-activity relationships of epigenetic targets using the activity landscape. Therefore, Sánchez-Cruz and Medina-Franco [16] pointed out the need for predictive models focused on epigenetic targets to assist drug research and development.

Machine learning algorithms have become very successful in multiple areas of drug discovery [17–19]. The development of deep learning algorithms has benefited various biochemical prediction areas [20]. A deep neural network was used to build a molecular property prediction model by extracting molecular descriptors–features [21, 22]. In fact, molecules are modeled as graphs where the graph size is arbitrary and its topology complex, and where their biological activity needs to be determined for drug discovery. Graph neural networks (GNN), a type of neural network structure that generates vector representation for nodes, have allowed investigating interatomic connections. Due to the very short forward propagation time of neural networks, the computation time is far shorter than that for density functional theory (DFT) as a widely used calculation method in fields such as physics and chemistry as well as those for other traditional quantum chemical calculation methods.

Because molecules are typical graph structures, atoms are regarded as nodes and chemical bonds as edges, and the prediction of molecular properties is transformed into a graph regression problem. Although GNNs have great advantages in processing graph structure data, the classical GNN model based on the abstract graph structure does not consider relevant laws of physics, which is fundamental for molecular property prediction. The gated GNN (GGNN) is one model for classical message passing based on gate recurrent-units (GRUs), which expands recurrent neural networks (RNNs) in a fixed T time step and uses back propagation through time (BPTT) to calculate the gradient, the recursive loop resulting in node representations [23]. Therefore, this kind of GNN can improve long-term information dissemination in graph structures and has a more powerful characterization ability.

In this article, we propose one integrated model by integrating GGNN-based feature extractions; various popular supervised classifiers are used to predict epigenetic targets though molecular property prediction such as support vector machine (SVM), random forest (RF) [24], artificial neural network (ANN) [25], and the conventional classifier extreme gradient boosting algorithm (XGBoost) [26]. The comparison demonstrated the advantage in targets prediction from our system versus former machine learning methods.

## 2. Materials and methods

### *2.1. Data set summary*

The proposed framework in this study was comprehensively evaluated using chemogenomic data for epigenetic targets. We extracted the quantitative compound-protein associations from ChEMBL 27 and PubChem [27] to build epigenetic target-associated compound datasets. We adopted the same 136 targets as Sessions et al. [28], where biological activity was reported as IC50, EC50, Ki, or Kd, and for each target document, compounds were labeled as active when they had unequivocally assigned activities $\leq$ 10 $\mu$M and as inactive in the opposite case. Then, we filtered the targets according to the following criteria: containing $\geq$ 30 biologically active compounds and $\geq$ 30 biologically inactive compounds. A total of 55 epigenetic targets, the same as in Sessions et al. [28], were analyzed: 16 kinases (KINs); 6 histone methyltransferases (HMTs); 3 histone acetyltransferases (HATs); DNA methyltransferase (DNMT1); 12 histone deacetylases (HDACs), 6 histone demethylases (HDMs), 2 proteins with dual activity (HDAC/HDM); 1 related to histone ubiquitination (USP7); 4 readers, such as one histone methyl-lysine binding protein (L3MBTL1); and other proteins, such as 3 chromatin remodelers (CHRs). These 55 compound datasets were consistent with those in Sessions et al. [28]. To facilitate the understanding of the purpose behind selecting these targets in this study, here is a brief explanation of each target:

- Kinases (KINs): Kinases are enzymes that modify proteins by adding phosphate groups through a process called phosphorylation. They play important roles in cellular signaling and regulating various cellular processes.
- Histone methyltransferases (HMTs): These enzymes add methyl groups to histone proteins, influencing gene expression and chromatin structure.
- Histone acetyltransferases (HATs): HATs add acetyl groups to histone proteins, leading to changes in gene expression by relaxing the chromatin structure.
- DNA methyltransferase (DNMT1): DNMT1 is responsible for adding methyl groups to DNA molecules, specifically to cytosine residues, which can affect gene expression.
- Histone deacetylases (HDACs): HDACs remove acetyl groups from histone proteins, resulting in gene repression and a more condensed chromatin structure.
- Histone demethylases (HDMs): HDMs remove methyl groups from histone proteins, influencing gene expression and chromatin remodeling.
- Proteins with dual activity (HDAC/HDM): These proteins possess both histone deacetylase (HDAC) and histone demethylase (HDM) activities, allowing them to modulate gene expression through multiple mechanisms.
- Histone ubiquitination-related protein (USP7): USP7 is involved in histone ubiquitination, which is the process of adding ubiquitin molecules to histone proteins. Histone ubiquitination can regulate gene expression and chromatin dynamics.
- Readers (such as L3MBTL1): Readers are proteins that recognize and bind to specific chemical modifications on histone proteins, such as methylated lysine residues. Lethal(3) malignant brain tumor-like protein 1 (L3MBTL1) is an example of a histone methyl-lysine binding protein.
- Chromatin remodelers (CHRs): These proteins modify the structure and organization of chromatin, allowing for changes in gene accessibility and regulation.

In the fields of chemical biology and drug research, a commonly used screening condition is $10\,\mu M$, which helps distinguish compounds with potential biological activity from inactive ones. A threshold of activity equal to or lower than $10\,\mu M$ is generally considered to indicate a certain level of biological activity. This choice of threshold is a common practice that aims to retain a reasonable number of active compounds while eliminating the majority of inactive ones.

By setting the criterion of having at least 30 biologically active compounds and 30 biologically inactive compounds for each target, we ensured a relatively balanced distribution of positive and negative samples in each target dataset. This balance was implemented to avoid any bias towards either class and to assist the model in learning and predicting effectively. Additionally, the choice of the number 30 is the minimum number of active and inactive compounds based on empirical methods. If the number is too small, it is not conducive to data balance and model training.

### *2.2. GGNN algorithm and machine learning methods*

The GGNN adopted an RNN principle based on GRU to perform the message passing in graph. The hidden state of one node is updated by its previous and adjacent hidden states:

$$h_v^{(t)} = \mathrm{GRU}(h_v^{(t-1)}, \sum_{\mu \in N(v)} W h_\mu^{(t-1)}),$$

where $h_v^{(t)}$ is the hidden state of node $v$ at time $t$, $N(v)$ is the set of neighbors of node $v$, and $W$ is the weight. The differences in propagation model between GGNN and GRU mainly include 1) the information transmission between different nodes is limited by the edges in the graph (the presence of edges, their direction and types) and 2) parameters in the propagation model, integrated updated door and reset door and the origin of name gated.

To use GGNN to learn the representation of molecular structures, each graph is composed of nodes and edges. The linear representation of molecular information was standardized by RDKit to process these simplified molecular input line systems (SMILES) encoding compounds; the adjacency matrix information can be obtained where the adjacency relationship between atoms is stored. On the other hand, a node is described by the atomic element names, elemental valence, elemental hydrogen connection numbers, atomic degree, and other properties.

SVM: For classification problems, SVM is a binary classification model based on a linear classifier defined in feature space with the maximum interval, SVM making it a substantially nonlinear classifier. The learning strategy of SVM is interval maximization, which can be formalized as a problem of solving convex quadratic program and equivalent to the minimization problem of regularized hinges loss function. SVM's learning algorithm is an optimization algorithm for solving convex quadratic programs.

Random Forest (RF): Random forest belongs to the bagging (bootstrap aggregation) method of ensemble learning, composed of many decision trees without correlations between different trees. When we perform the classification task, new input samples will enter, and each decision tree in the forest will be allowed to judge and classify separately. Each decision tree will get its own classification result. The random forest will use the result having most classifications as final result.

ANN: A neural network is a wide network of parallel interconnections composed of adaptive simple units. Its organization can simulate the interaction of biological nervous system to real word objects. ANN is composed of multiple neuron models connected according to certain rules. Its classification algorithm belongs to supervised learning algorithm.

XGBoost [29]: XGBoost (eXtreme Gradient Boosting) is a boosting algorithm, a lifting tree model that integrates many trees. It grows a tree continuously through feature splitting, to learn a tree in each around; then, it fits the residual between the predicted value and the actual value of the previous round. For a given dataset $D = \{(x_i, y_i)\}$, XGBoost uses additive training to learn $K$ trees, via advanced regularization as in the following formula:

$$\tilde{y} = \phi(x_i) = \sum_{k=1}^{K} f_k(x_i), \qquad f_k \in \Gamma,$$

where $\Gamma$ is the hypothesis space of all decision trees, and $f_k$ is the tree that needs to be learned during training. Each round of training a tree is designed to minimize the loss function, which not only measures the fitting error of the model but also adds a regularization term to limit tree complexity and prevents over-fitting by the penalty term of the complexity of each tree:

$$L(\phi) = \sum_i l(\tilde{y}_i, y_i) + \sum_k \Omega(f_k),$$

where $\Omega(f) = \gamma T + \frac{1}{2}\lambda\|\omega\|^2$.

The supervised learner XGBoost is a popular and very efficient machine learning algorithm for prediction problems in various biochemical fields.

### *2.3. Prediction*

This study followed three major steps, as shown in Figure 1. First, the raw molecular data was formatted as a graph structure using a tensor that stored atom properties by method similar to those in [30]. Second, we used the GGNN to learn the graph representations from the molecular features. Finally, we loaded these graph representations as sample features into the four classifiers: SVM, RF, ANN, and XGBoost.

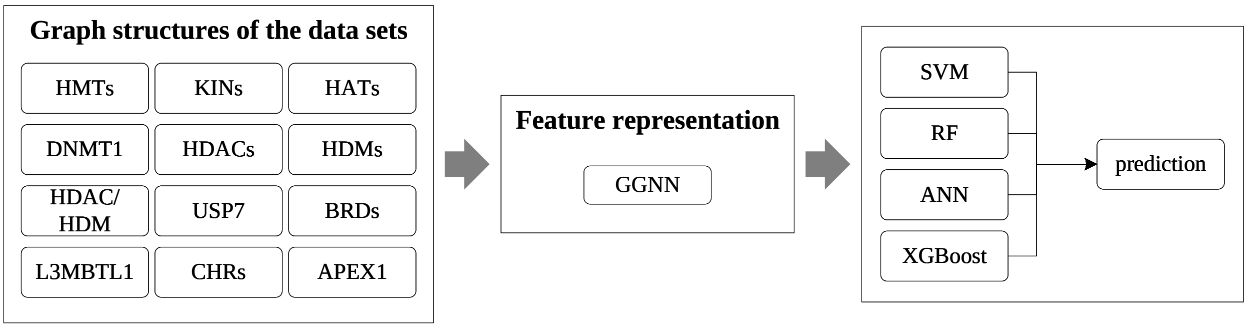


**Figure 1.** The overall workflow of building an efficient and accurate predictor for epigenetic targets.

### *2.4. Performance metrics*

This study evaluated the proposed GGNN-based feature extraction and target prediction. Their respective performance metrics are defined as follows:

This binary classification problem may be evaluated using the metrics specificity (Sp), sensitivity (Sn), accuracy (ACC), balanced accuracy (BA), Matthews correlation coefficient (MCC), and true positive rate (TPR). The numbers of true positives and false negatives are defined as TP and FN. Moreover, F1, the harmonic mean, was adopted to describe the accuracy in samples by balancing the precision and TPR.

$$
\begin{aligned}
\mathrm{ACC} &= (\mathrm{TP}+\mathrm{TN})/(\mathrm{TP}+\mathrm{FN}+\mathrm{TN}+\mathrm{FP}),\\
\mathrm{BA} &= 0.5\mathrm{TP}/(\mathrm{TP}+\mathrm{FN})+0.5\mathrm{TN}/(\mathrm{TN}+\mathrm{FP}),\\
\mathrm{MCC} &= (\mathrm{TP}\times\mathrm{TN}-\mathrm{FP}\times\mathrm{FN})\\
&\quad (\mathrm{TP}+\mathrm{FP})^{-\frac{1}{2}}(\mathrm{TP}+\mathrm{FN})^{\frac{1}{2}}\\
&\quad (\mathrm{TN}+\mathrm{FP})^{-\frac{1}{2}}(\mathrm{TN}+\mathrm{FN})^{-\frac{1}{2}},\\
\mathrm{F1} &= 2/(\frac{1}{\mathrm{PRECISION}}+\frac{1}{\mathrm{TPR}}).
\end{aligned}
$$

### *2.5. Parameter tuning*

For every dataset of these 55 epigenetic targets, the molecular map representation and activity classification information were obtained through pre-processing, as shown in Figure 2. As for the pre-process, we first filtered the original 136 targets. We collected all the quantitative compound-protein associations from single-protein assays related to the 136 epigenetic targets, where biological activity was reported as IC50, EC50, Ki, or Kd. For each target document, compounds were labeled as active when they had unequivocally assigned active $\leq$ 10 $\mu M$ and as inactive in the opposite case. Then, we filtered the targets according to the following criteria: containing $\geq$ 30 biologically active compounds and $\geq$ 30 biologically inactive compounds. The open-source cheminformatics toolkit RDKit, version 2020.03.1, was adopted to curate these compounds. Then, by the same process as [16], we also extended compounds with annotated quantitative data (IC50) from PubChem (accessed October 2020), finally resulting in 55 targets with at least 30 active and inactive compounds. Using the classification information in the original datasets as data labels and the temporary representation of molecular graphs and the adjacency matrix as input graph information, the GGNN was used to train EPOCH = 300 rounds, which is enough to make the model converge. Some loss curves of target proteins are shown in Figure 3.

### *2.6. Experimental environment*

The experiment was run on Ubuntu SMP x86_64 x86_64 X86_64 GNU/Linux with Linux bit 4.15.0-173-generic #182-Ubuntu SMP x86_64 X86_64 GNU/Linux. The running environment was constructed using Conda. Python 3.8.13 was used as programming language, using an NVIDIA GeForce RTX 3090 graphics card, the RDKit 2022.3.1 package for molecular fingerprint processing, and Pytorch 1.11.0 was used to build the deep learning network. A CSV library was used for data processing, and the Matplotlib library was used for data visualization.

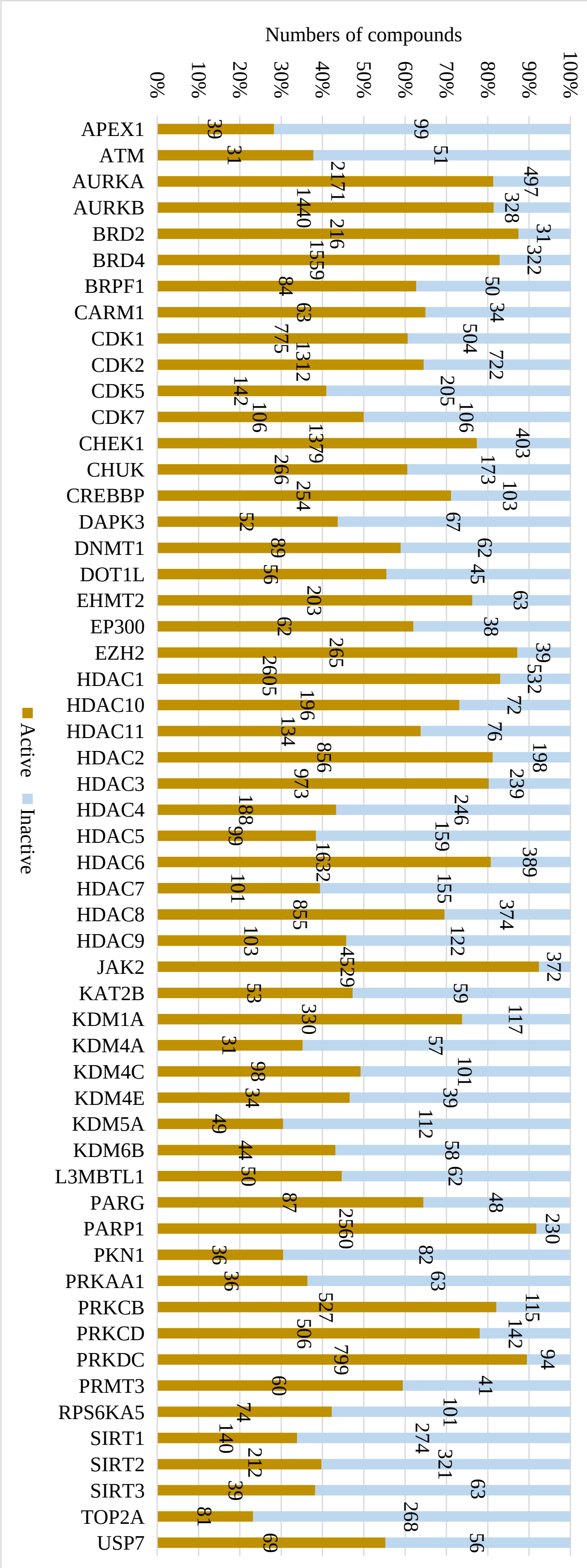

**Figure 2.** Distribution of the epigenetic targets included in this work. The x-axis represents targets proteins, and the y-axis represents numbers of compounds related with these proteins.

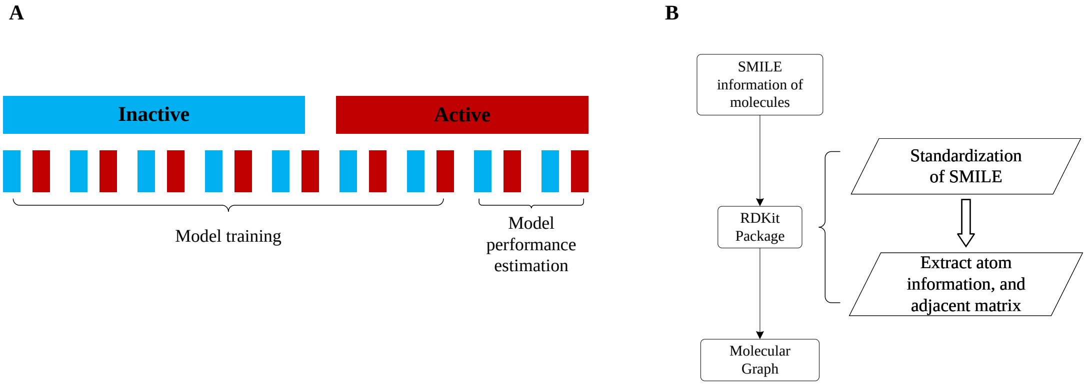


**Figure 3.** The validation strategies employed for epigenetic target prediction in this article. Blue and red boxes represent inactive and active compounds related to the target protein separately.

## 3. Results

### *3.1. Epigenetic target prediction*

The size and composition of target-associated compounds and the average proportion of active compounds of 59.3% is shown in Figure 4. The predictive model for epigenetic target prediction built (Figure 2) contained 55 single-target binary classification tasks, and involved the performance comparison of four different classifiers based on GGNN network feature extraction. The active represented compounds were labeled as such when they had biological activities (such as IC50, EC50, Ki, or Kd) $\leq$ 10 $\mu$M and as inactive in the opposite case. Network features were derived from the GGNN.

### *3.2. Comparison with the original machine learning features*

To compare our generated models of feature extraction in a more global context, the cross-validated predictions for our model with the original Fingerprint algorithm Molecular ACCess System (MACCS) [31] were stored and used to compute various performance metrics (BA, F1, and MCC). A summary of the models' performance is presented in Table 1. We also employed a stratified 10-fold cross-validation approach for 55 single-target binary prediction tasks. The dataset was divided into training, test, and validation sets with proportions of 70%, 20%, and 10%, respectively. The name and best value trained by 10-fold cross validation is shown in Table 2. In KNN classification, we utilized the Jaccard distance as the metric to measure the distance between data points. The optimal number of nearest neighbors was selected from the set of values: 1, 3, 5, 7, and 10. For RF classification, we employed an ensemble of decision trees. Each decision tree predicted a label for a given sample, and the final prediction for the sample was determined by the majority vote among the tree predictions. We used a fixed number of 1000 decision trees in the ensemble. SVM classification aimed to construct the best separation between the two classes by maximizing the distance between instances that belonged to different classes. A cost hyperparameter, which controlled the trade-off between errors, was optimized. We considered the following candidate values for the cost parameter:

0.01, 0.1, 1, 10, and 100. For XGBoost classification, we tuned three parameters: the learning rate, maximum depth, and minimum child weight. We considered five candidate values for each parameter: 0.01, 0.05, 0.1, 0.15, and 0.2.

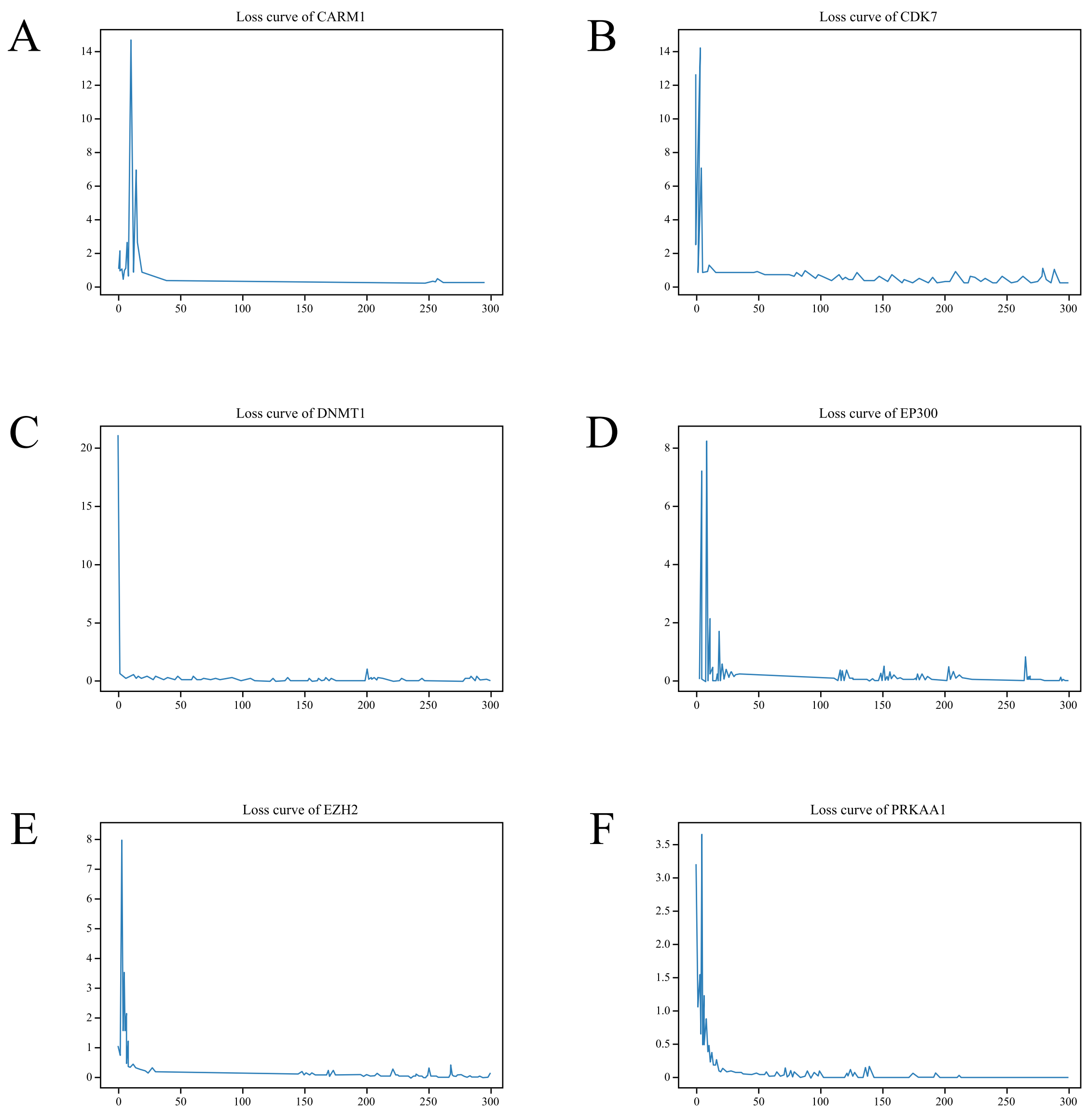


**Figure 4.** The parameters training in GGNN for 300 epoches. The loss curves of six target proteins were evaluated. The horizontal axis gives the epoches.

**Remark 1.** *We adopt a nested cross validation strategy. The outer loop divides the data into a training set and a testing set. In the inner loop, the training set is further optimized for hyperparameters through 10 fold cross validation. All performance metrics reported are based on the calculation results of an independent test set, which has not participated in model training or parameter tuning at any stage.*

We chose MACCS+KNN, MACCS+XGBoost, ECFP+KNN, and Mol2Vec+XGBoost as baselines. All methods were evaluated on 55 epigenetic targets. Based on the performance of GGNN+XGBoost, we divided the 55 targets into two categories. Among them, 42 targets were named as the H category, which means their performance in this category is higher than other baselines. The other 13 targets were labeled as the L category, which means their performance is slightly lower than other baselines. The average performance of each model in the dominant and disadvantaged categories (ACC, BA, MCC) is given to better present the performance comparison of the models. The results are shown in Table 1.

**Table 1.** The average performance of different baselines among 55 targets.

| Models | ACC (H) | BA (H) | MCC (H) | ACC (L) | BA (L) | MCC (L) |
|---|---|---|---|---|---|---|
| GGNN+XGBoost | 0.830 | 0.810 | 0.640 | 0.780 | 0.764 | 0.575 |
| ECFP+KNN | 0.780 | 0.770 | 0.595 | 0.805 | 0.773 | 0.585 |
| Mol2Vec+XGBoost | 0.794 | 0.763 | 0.614 | 0.814 | 0.783 | 0.572 |
| MACCS+KNN | 0.778 | 0.756 | 0.563 | 0.798 | 0.768 | 0.581 |

**Table 2.** The hyperparameter selection for the model.

| GGNN+XGBoost | GGNN+SVM | GGNN+KNN | GGNN+RF |
|---|---|---|---|
| max_depth = 6 | kernel = rbf | n_neighbors = 5 | n_estimators = 100 |
| min_child_weight = 4 | C = 1 | weights = uniform | criterion = gini |
| XGBoost_learning_rate = 0.001 | gamma = auto | algorithm = auto | - |
| GGNN_learning_rate = 0.004 | - | - | - |
| GGNN_epoch = 200 | GGNN_epoch = 200 | GGNN_epoch = 200 | GGNN_epoch = 200 |
| GGNN_batch = 64 | GGNN_batch = 64 | GGNN_batch = 64 | GGNN_batch = 64 |

### *3.3. Evaluation of different supervised classifiers*

Using ACC as performance metric to show the results of these 55 single targets (Figure 5), we detected that XGBoost had the highest ACC for 27 out of the 55 targets, making it the best choice; meanwhile, SVM had 9 out of 55, RF had 12 out of 55, and RNN had 7 out of 55. Particularly, XGBoost achieved the best accuracy for targets HDAC2, HDAC8, and HDAC11, belonging to the HDAC family-the main targets of early epigenetic drug development and targets RKN1, PRKCD, belonging to kinase (KIN) family, one novel epigenetic target that proved be therapeutic options for advanced prostate cancer treatment [32]. However, in bromodomain (BRD) targets such as BRD2 or BRD4, RF was the best choice. Meanwhile, the cofactor involved in DNA demethylation APEX1 went with SVM. Even if RF, KNN, and SVM achieved the highest accuracy in some datasets, the unstable distribution of results makes it difficult to reflect the performance of these classifiers. Table 3 shows the average accuracy in these 55 target datasets for different classifiers.

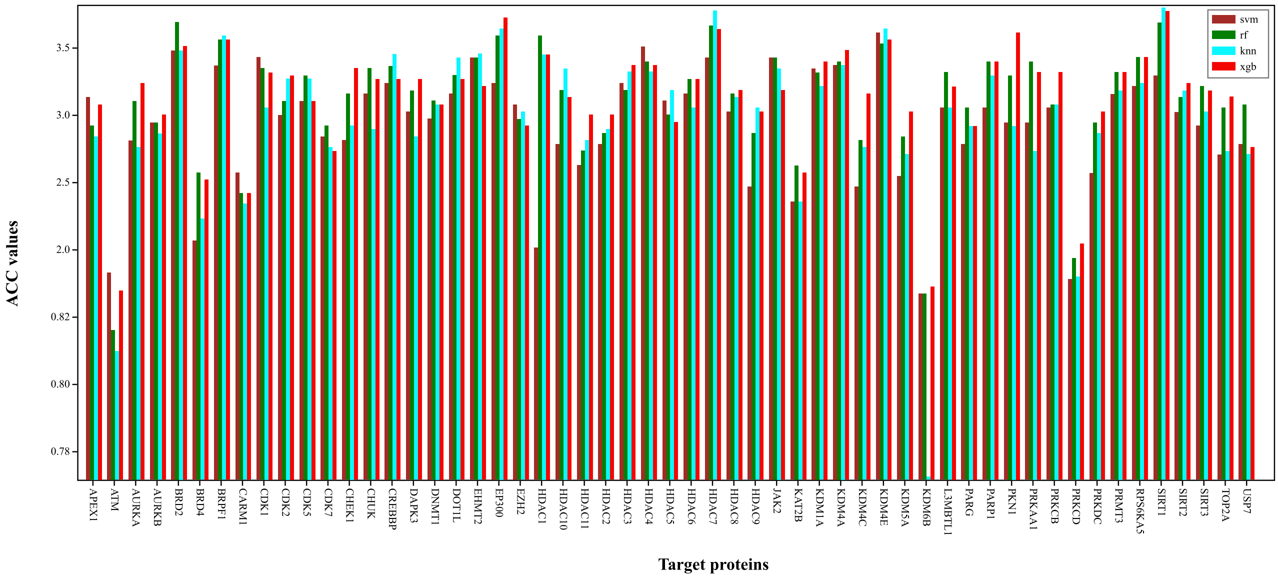


**Figure 5.** The ACC performance results of 55 target datasets with different classifiers. The x-axis shows different target proteins, and the y-axis measures the ACC values.

**Table 3.** Average accuracy of different classifiers based on GGNN in 55 datasets.

| Classifier | $\overline{\text{ACC}}$ | Standard deviation $\sigma$ |
|---|---|---|
| SVM | 0.811387 | 0.09711 |
| RF | 0.811376 | 0.10238 |
| KNN | 0.813017 | 0.09129 |
| XGBoost | 0.815841 | 0.09122 |

**Remark 2.** *Our method is the application of one novel graph neural network for feature extraction in prediction task of epigenetic targets. To explain the performance of our generated model in feature extraction, we compared five machine learning methods (RF, gradient boost trees, SVM, FFNN, and kNN) used by Sánchez-Cruz and Medina-Franco [16]. While Sánchez-Cruz and Medina-Franco performed multi-target validation by combining the predictions from the 55 individual models, our research primarily focused on evaluating the single-target performance of our GGNN model. However, it is important to note that the multi-target validation employed by Sánchez-Cruz and Medina-Franco can be seen as another manifestation of performance evaluation for single-target predictions.*

To rigorously evaluate the model's generalization ability (i.e., scaffold jumping ability) toward novel chemical structures, this study employs a data segmentation method based on the Bemis-Murcko molecular skeleton. To begin, extract the Bemis-Murcko skeleton of all compounds using RDKit. Subsequently, ensure that all molecular skeletons in the test set do not appear in the training set. This segmentation method aims to simulate real drug discovery scenarios, predicting the activity of lead compounds with completely new structures in order to verify whether the model has learned potential structure-activity relationships rather than simply memorizing known structural patterns. The results are presented in Table 4. In more challenging skeleton segmentation scenarios, the performance degradation of the GGNN method is significantly smaller than that of all traditional molecular representation based methods (ECFP, MACCS, and Mol2Vec), indicating that GGNN has learned more universal chemical features from molecular graph structures.

**Table 4.** Performance Average under Different criteria in 55 datasets.

| Model | Segmentation method | ACC | BA | MCC |
|---|---|---|---|---|
| GGNN+XGBoost | Random | 0.8158 | 0.7962 | 0.6305 |
| GGNN+XGBoost | Skeleton | 0.7894 | 0.7660 | 0.5920 |
| ECFP+XGBoost | Random | 0.8012 | 0.7825 | 0.6018 |
| ECFP+XGBoost | Skeleton | 0.7246 | 0.7034 | 0.4823 |
| MACCS+XGBoost | Random | 0.7945 | 0.7739 | 0.5874 |
| MACCS+XGBoost | Skeleton | 0.7128 | 0.6915 | 0.4637 |
| Mol2Vec+XGBoost | Random | 0.7893 | 0.7681 | 0.5796 |
| Mol2Vec+XGBoost | Skeleton | 0.7189 | 0.6972 | 0.4715 |

Table 5 shows the prediction results of the target Janus kinase 2 (JAK2), in which compound molecules matched the greatest number of samples. On the other hand, Table 6 shows the results for the target lysine demethylase 4E (KDM4E), in which compound molecules matched the smallest number of samples. Tables 7–9 show the results of targets from smallest to largest. The number of samples influenced the model's classification effect. The accuracy of the features extracted from KDM4E for activity classification is only 81.82%, while the classification model accuracy increases with increasing data volume, reaching 95.7% for JAK2, the dataset with the highest data volume. Simultaneously, the index F1, which measures the difference of sample data ignored by the model, also increases from the initial 0.8 to 0.9749. Due to data limitations, GGNN extracted features are very limited, and it is difficult to learn the feature representation belonging to the current data set, resulting in insufficient effectiveness of feature extraction and poor classification effect. However, in large data sets, the number of molecules and differences between molecular features enables GGNN to extract similarities and differences between molecules to obtain compounds which can represent the current gene activity under feature classification, and the traditional classifier according to SDAID features can be correctly classified.

**Table 5.** Gene JAK2 comparative results of different classifiers.

| Evaluation Index | RF | SVM | KNN | XGBoost |
|---|---|---|---|---|
| ACC | 0.9524 | 0.9490 | 0.9483 | 0.9531 |
| PRECISION | 0.9570 | 0.9518 | 0.9556 | 0.9629 |
| BA | 0.6977 | 0.6596 | 0.6864 | 0.7388 |
| F1 | 0.9749 | 0.9732 | 0.9727 | 0.9751 |
| MCC | 0.5543 | 0.5046 | 0.5129 | 0.5821 |
| Numbers of samples | 1470 | 1470 | 1470 | 1470 |

**Table 6.** Gene KDM4E comparative results of different classifiers.

| Evaluation Index | RF | SVM | KNN | XGBoost |
|---|---|---|---|---|
| ACC | 0.8182 | 0.7727 | 0.8182 | 0.8182 |
| PRECISION | 0.8889 | 0.875 | 0.8889 | 0.8889 |
| BA | 0.8182 | 0.7727 | 0.8182 | 0.8182 |
| F1 | 0.8 | 0.7368 | 0.8 | 0.8 |
| MCC | 0.6472 | 0.5669 | 0.6472 | 0.6472 |
| Numbers of samples | 22 | 22 | 22 | 22 |

**Table 7.** Gene KDM1A comparative results of different classifiers.

| Evaluation Index | RF | SVM | KNN | XGBoost |
|---|---|---|---|---|
| ACC | 0.8881 | 0.8881 | 0.8881 | 0.8731 |
| PRECISION | 0.9468 | 0.9468 | 0.9468 | 0.9457 |
| BA | 0.8680 | 0.8781 | 0.8781 | 0.8781 |
| F1 | 0.9223 | 0.9223 | 0.9223 | 0.9110 |
| MCC | 0.7259 | 0.7259 | 0.7259 | 0.6969 |
| Numbers of samples | 134 | 134 | 134 | 134 |

**Table 8.** Gene HDAC8 comparative results of different classifiers.

| Evaluation Index | RF | SVM | KNN | XGBoost |
|---|---|---|---|---|
| ACC | 0.9133 | 0.9051 | 0.9079 | 0.9051 |
| PRECISION | 0.9163 | 0.8984 | 0.9053 | 0.9118 |
| BA | 0.9016 | 0.8878 | 0.8929 | 0.8936 |
| F1 | 0.9319 | 0.9266 | 0.9283 | 0.9254 |
| MCC | 0.8135 | 0.7960 | 0.8017 | 0.7960 |
| Numbers of samples | 369 | 369 | 369 | 369 |

**Table 9.** Gene HDAC1 comparative results of different classifiers.

| Evaluation Index | RF | SVM | KNN | XGBoost |
|---|---|---|---|---|
| ACC | 0.8895 | 0.8895 | 0.8842 | 0.8980 |
| PRECISION | 0.9046 | 0.8999 | 0.9021 | 0.9094 |
| BA | 0.7315 | 0.7189 | 0.7233 | 0.7467 |
| F1 | 0.9359 | 0.9363 | 0.9328 | 0.9408 |
| MCC | 0.5567 | 0.5509 | 0.5346 | 0.55933 |
| Numbers of samples | 941 | 941 | 941 | 941 |

## 4. Discussion and conclusions

The role of epigenetic mechanisms in cancer development is well studied. Both DNA hypomethylation and hypermethylation have been associated with tumor growth. Low and high acetylation of histone H3 and H4 are involved in tumor suppression; this effect is independent of altered DNA methylation. In addition to cancer, epigenetic factors are also involved in inflammation, autoimmune, metabolic, and blood diseases. For example, haploid deletion of HDAC4 can cause intellectual development disorders and short toes. Abnormalities in other epigenetic mechanisms have also been associated with multiple intellectual development disorders. Some epigenetic abnormalities are inherited, but more often they are acquired, which is important for drug development because it indicates that the epigenetic state is reversible and can be affected by drugs. Not many epigenetic drugs have reached late-stage clinical development, however, because of the complex and expensive research processes. This requires artificial intelligence or data mining methods to help predict epigenetic targets.

In this study, we proposed the extraction of GNN-based features of molecular properties. Then, we integrated four supervised classifiers for prediction of epigenetic targets. A comprehensive evaluation was conducted to demonstrate that the feature extraction ability of original machine learning methods may be improved by GNN and different supervised classifiers. The number of molecules and differences between molecular structures allowed the GGNN to extract molecular similarities and differences among molecules to obtain compounds that can represent the current gene activity under the feature classification. The proposed framework may efficiently improve prediction of epigenetic targets.

**Table 10.** The performance results of 55 target datasets with the GGNN + XGBoost model.

| Target | MCC | F1 | Precision | Recall | Accuracy |
|---|---|---|---|---|---|
| APEX1 | 0.306904851 | 0.920676203 | 0.864468864 | 0.984700974 | 0.857309942 |
| ATM | 0.282828283 | 0.555555556 | 0.555555556 | 0.555555556 | 0.661971831 |
| AURKA | 0.682470014 | 0.948051948 | 0.939914163 | 0.956331878 | 0.913043478 |
| AURKB | 0.637410304 | 0.941810345 | 0.929787234 | 0.954148472 | 0.902173913 |
| BRD2 | 0.163203056 | 0.425531915 | 0.526315789 | 0.357142857 | 0.61971831 |
| BRD4 | -0.034126813 | 0.430769231 | 0.378378378 | 0.5 | 0.478873239 |
| BRPF1 | 0.482788134 | 0.714285714 | 0.727272727 | 0.701754386 | 0.744 |
| CARM1 | 0.344155244 | 0.769230769 | 0.681818182 | 0.882352941 | 0.689655172 |
| CDK1 | 0.593209891 | 0.926739927 | 0.906810036 | 0.947565543 | 0.879518072 |
| CDK2 | 0.08159656 | 0.652173913 | 0.731707317 | 0.588235294 | 0.561643836 |
| CDK5 | 0.466186471 | 0.922024624 | 0.907133244 | 0.937413074 | 0.866666667 |
| CDK7 | 0.306904851 | 0.920676203 | 0.864468864 | 0.984700974 | 0.857309942 |
| CHEK1 | 0.461710971 | 0.926500337 | 0.89921466 | 0.955493741 | 0.87251462 |
| CHUK | 0.248792328 | 0.897637795 | 0.826086957 | 0.982758621 | 0.819444444 |
| CREBBP | 0.368348682 | 0.822695035 | 0.828571429 | 0.816901408 | 0.744897959 |
| DAPK3 | 0.593209891 | 0.926739927 | 0.906810036 | 0.947565543 | 0.879518072 |
| DNMT1 | 0.548795472 | 0.830188679 | 0.88 | 0.785714286 | 0.785714286 |
| DOT1L | 0.24743583 | 0.685714286 | 0.571428571 | 0.857142857 | 0.607142857 |
| EHMT2 | -0.05830702 | 0.88372093 | 0.802816901 | 0.982758621 | 0.791666667 |
| EP300 | 0.560134283 | 0.921389397 | 0.9 | 0.943820225 | 0.870481928 |
| EZH2 | 0.402620115 | 0.942675159 | 0.891566265 | 1 | 0.894117647 |

*Continued on next page*

| Target | MCC | F1 | Precision | Recall | Accuracy |
|---|---|---|---|---|---|
| HDAC1 | 0.56812764 | 0.938941655 | 0.916556291 | 0.962447844 | 0.894736842 |
| HDAC10 | 0.438355035 | 0.854545455 | 0.796610169 | 0.921568627 | 0.780821918 |
| HDAC11 | 0.344265186 | 0.794871795 | 0.738095238 | 0.861111111 | 0.714285714 |
| HDAC2 | 0.475607723 | 0.917171717 | 0.876447876 | 0.961864407 | 0.859106529 |
| HDAC3 | 0.593209891 | 0.926739927 | 0.906810036 | 0.947565543 | 0.879518072 |
| HDAC4 | 0.596724519 | 0.76635514 | 0.82 | 0.719298246 | 0.8 |
| HDAC5 | 0.245525819 | 0.588235294 | 0.5 | 0.714285714 | 0.605633803 |
| HDAC6 | 0.626550798 | 0.939524838 | 0.929487179 | 0.949781659 | 0.898550725 |
| HDAC7 | 0.436464614 | 0.654545455 | 0.642857143 | 0.666666667 | 0.732394366 |
| HDAC8 | 0.697718516 | 0.905172414 | 0.867768595 | 0.945945946 | 0.868263473 |
| HDAC9 | 0.482381911 | 0.783783784 | 0.743589744 | 0.828571429 | 0.746031746 |
| JAK2 | 0.345341314 | 0.880294659 | 0.865942029 | 0.895131086 | 0.804216867 |
| KAT2B | 0.266966173 | 0.647058824 | 0.578947368 | 0.733333333 | 0.625 |
| KDM1A | 0.558290526 | 0.903553299 | 0.898989899 | 0.908163265 | 0.849206349 |
| KDM4A | 0.3 | 0.222222222 | 1 | 0.125 | 0.730769231 |
| KDM4C | 0.439515621 | 0.680851064 | 0.64 | 0.727272727 | 0.722222222 |
| KDM4E | 0.291755818 | 0.428571429 | 0.75 | 0.3 | 0.652173913 |
| KDM5A | 0.434440897 | 0.526315789 | 0.714285714 | 0.416666667 | 0.8 |
| KDM6B | 0.276394996 | 0.52173913 | 0.666666667 | 0.428571429 | 0.64516129 |
| L3MBTL1 | -0.021744269 | 0.722891566 | 0.638297872 | 0.833333333 | 0.589285714 |
| PARG | 0.475607723 | 0.917171717 | 0.876447876 | 0.961864407 | 0.859106529 |
| PARP1 | 0.269430126 | 0.732394366 | 0.742857143 | 0.722222222 | 0.660714286 |
| PKN1 | 0.344265186 | 0.794871795 | 0.738095238 | 0.861111111 | 0.714285714 |
| PRKAA1 | -0.038461617 | 0.733944954 | 0.689655172 | 0.784313725 | 0.602739726 |
| PRKCB | 0.614991223 | 0.936877076 | 0.933774834 | 0.94 | 0.894444444 |
| PRKCD | 0.438355035 | 0.854545455 | 0.796610169 | 0.921568627 | 0.780821918 |
| PRKDC | 0.056295069 | 0.793103448 | 0.707692308 | 0.901960784 | 0.671232877 |
| PRMT3 | 0.666666667 | 0.864864865 | 0.761904762 | 1 | 0.821428571 |
| RPS6KA5 | 0.314038633 | 0.900793651 | 0.847014925 | 0.961864407 | 0.828178694 |
| SIRT1 | 0.667367616 | 0.780487805 | 0.744186047 | 0.820512821 | 0.848739496 |
| SIRT2 | 0.331820185 | 0.557692308 | 0.644444444 | 0.491525424 | 0.689189189 |
| SIRT3 | 0.666722895 | 0.8 | 0.833333333 | 0.769230769 | 0.838709677 |
| TOP2A | 0.444170697 | 0.916666667 | 0.861940299 | 0.978813559 | 0.855670103 |
| USP7 | 0.333623062 | 0.804878049 | 0.717391304 | 0.916666667 | 0.714285714 |

Our proposed method offers several advantages for predicting epigenetic targets: First, by integrating the GGNN, we effectively capture the structural information and relationships between molecules, resulting in a more accurate representation of the data. Second, our method utilizes the GGNN to extract informative features from the molecules, enabling the capture of relevant characteristics for predicting epigenetic targets. The performance results of 55 target datasets with the GGNN+XGBoost model is shown as Table 10. The GGNN+XGBoost model employed in this study utilizes the GGNN to learn graph representations from molecular features and then feeds these representations into the XGBoost classifier for classification or prediction tasks. This combined approach leverages the strengths of both neural networks and gradient boosting algorithms to achieve improved performance in molecular data analysis. This highlights the effectiveness of our approach in predicting epigenetic targets. Then, the superiority of the GGNN is verified by comparing three integrated predictive models: GCN+XGBoost, GGNN+XGBoost, and DMPNN+XGBoost, as illustrated in Table 11. It can be known that the effect of the GGNN+XGBoost model is better.

**Table 11.** The performance for three integrated predictive models.

| Evaluation Index | GCN | GGNN | DMPNN | GCN (CW) | GGNN (CW) | DMPNN (CW) |
|---|---|---|---|---|---|---|
| ACC | 0.7557 | 0.8325 | 0.8093 | 0.7274 | 0.8148 | 0.7732 |
| BA | 0.3229 | 0.5055 | 0.4650 | 0.6259 | 0.7771 | 0.7038 |
| MCC | 0.1813 | 0.3940 | 0.3910 | 0.4944 | 0.6202 | 0.5730 |

To address the issue of class imbalance, we introduce weight adjustment based on class frequency in the training process of all models to ensure that a few classes (active compounds) receive appropriate weights in the loss function. Specifically, the weight $w_i$ of each category $i$ is calculated as $w_i = \frac{N}{KN_i}$, where $N$ is the total number of samples, $K$ is the number of categories, and $N_i$ is the number of samples in category $i$. This weight ensures that the model appropriately focuses on minority group categories during training, thereby improving its performance in identifying rare categories. Table 11 shows the improved results. Receiver-operating characteristic (ROC) curves and area under the curve (AUC) models and precision-recall curves are shown in Figures 6 and 7.

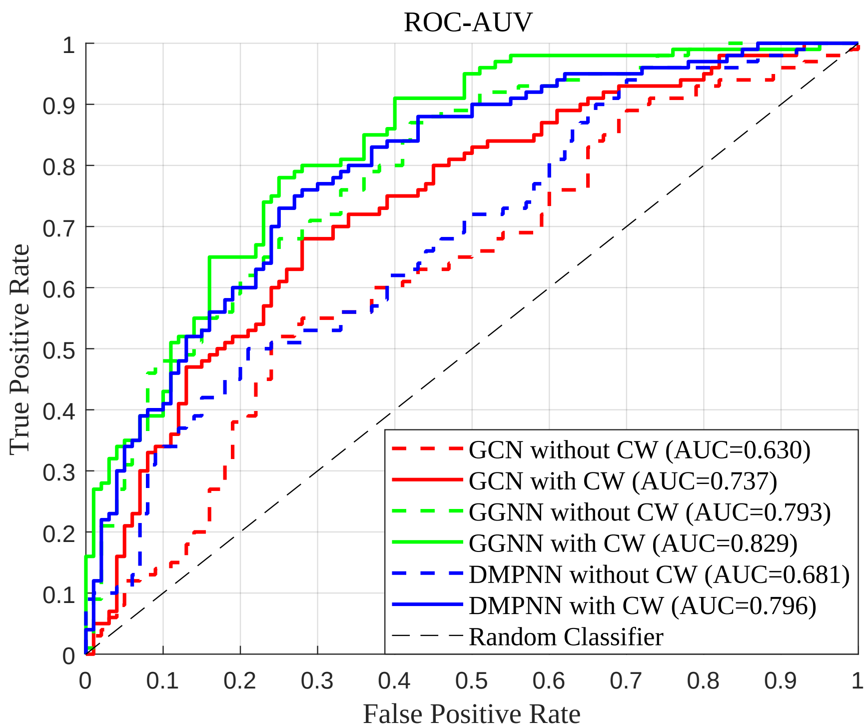


**Figure 6.** Receiver operating characteristic and area under the curve of models.

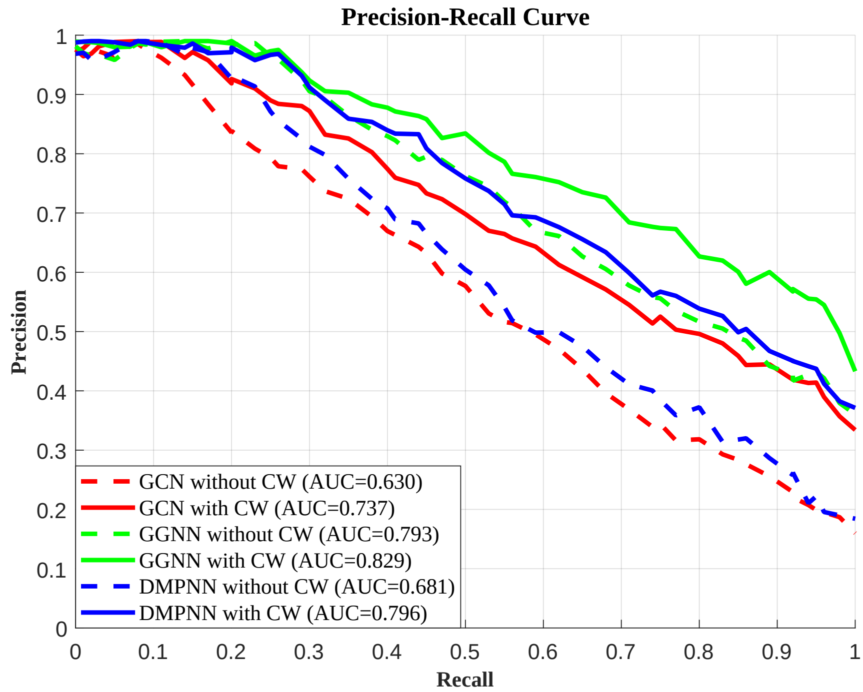


**Figure 7.** Precision-recall curves of models.

However, it is important to consider the limitations of our method. Firstly, the predictive performance is influenced by the availability and quality of the screened chemogenomic datasets, including their size, diversity, and potential biases. Secondly, while our model achieves high accuracy, the interpretability of its decisions may be limited.

Epigenetic regulation is influenced by various factors, including DNA methylation, histone modifications, and non-coding RNA. Integrating multi-omics data, such as transcriptomics and proteomics, can provide a more comprehensive understanding of the epigenetic landscape. Future research should explore methods that can effectively integrate and analyze these diverse data types, and we are committed to applying the GGNN+XGBoost model to multi-objective classification on different datasets.

## Data statement

The datasets generated and/or analysed during the current study are available in ChEMBL and PubChem.
https://www.ebi.ac.uk/chembl/
https://pubchem.ncbi.nlm.nih.gov/

## Use of AI tools declaration

The authors declare they have not used Artificial Intelligence (AI) tools in the creation of this article.

## Conflict of interest

The authors declare there are no conflicts of interest.

## References


1. C. H. Waddington, The epigenotype, *Int. J. Epidemiol.*, **41** (2012), 10–13. https://doi.org/10.1093/ije/dyr184
2. M. Esteller, Molecular origins of cancer: Epigenetics in cancer, *N. Engl. J. Med.*, **358** (2008), 1148–1159. https://doi.org/10.1056/NEJMra072067
3. V. Januar, R. Saffery, J. Ryan, Epigenetics and depressive disorders: A review of current progress and future directions, *Int. J. Epidemiol.*, **44** (2015), 1364–1387. https://doi.org/10.1093/ije/dyu273
4. C. Küçükali, M. Kürtüncü, A. Çoban, M. Çebi, E. Tüzün, Epigenetics of multiple sclerosis: An updated review, *NeuroMol. Med.*, **17** (2015), 83–96. https://doi.org/10.1007/s12017-014-8298-6
5. M. Brindisi, A. Saraswati, S. Brogi, S.Gemma, S. Butini, G. Campiani, Old but gold: Tracking the new guise of histone deacetylase 6 (HDAC6) enzyme as a biomarker and therapeutic target in rare diseases, *J. Med. Chem.*, **63** (2019), 23–39. https://doi.org/10.1021/acs.jmedchem.9b00924
6. A. Lera, A. Ganesan, Two-hit wonders: The expanding universe of multitargeting epigenetic agents, *Curr. Opin. Chem. Biol.*, **57** (2020), 135–154. https://doi.org/10.1016/j.cbpa.2020.05.009
7. U. Norinder, J. Naveja, E. López-López, D. Mucs, J. L. Medina-Franco, Conformal prediction of HDAC inhibitors, *SAR QSAR Environ. Res.*, **30** (2019), 265–277. https://doi.org/10.1080/1062936X.2019.1591503
8. P. Filippakopoulos, J. Qi, S. Picaud, Y. Shen, W. B. Smith, O. Fedorov, et al., Selective inhibition of BET bromodomains, *Nature*, **468** (2010), 1067–1073. https://doi.org/10.1038/nature09504
9. C. Allis, T. Jenuwein, The molecular hallmarks of epigenetic control, *Nat. Rev. Genet.*, **17** (2016), 487–500. https://doi.org/10.1038/nrg.2016.59
10. K. Cao, G. Wang, W. Li, L. Zhang, R. Wang, Y. Huang, et al., Histone deacetylase inhibitors prevent activation-induced cell death and promote anti-tumor immunity, *Oncogene*, **34** (2015), 5960–5970. https://doi.org/10.1038/onc.2015.46
11. M. Glozak, N. Sengupta, X. Zhang, E. Seto, Acetylation and deacetylation of non-histone proteins, *Gene*, **363** (2005), 15–23. https://doi.org/10.1016/j.gene.2005.09.010
12. J. Naveja, J. Medina-Franco, Insights from pharmacological similarity of epigenetic targets in epipolypharmacology, *Drug Discovery Today*, **23** (2018), 141–150. https://doi.org/10.1016/j.drudis.2017.10.006
13. J. Nanda, R. Kumar, G. Raghava, dbEM: A database of epigenetic modifiers curated from cancerous and normal genomes, *Sci. Rep.*, **6** (2016), 19340. https://doi.org/10.1038/srep19340
14. Y. Medvedeva, A. Lennartsson, R. Ehsani, I. V. Kulakovskiy, I. E. Vorontsov, P. Panahandeh, et al., EpiFactors: A comprehensive database of human epigenetic factors and complexes, *Database (Oxford)*, **2015** (2015), bav067. https://doi.org/10.1093/database/bav067

15. S. Loharch, I. Bhutani, K. Jain, P. Gupta, D. K. Sahoo, R. Parkesh, EpiDBase: A manually curated database for small molecule modulators of epigenetic landscape, *Database (Oxford)*, **2015** (2015), bav013. https://doi.org/10.1093/database/bav013

16. N. Sánchez-Cruz, J. Medina-Franco, Epigenetic target fishing with accurate machine learning models, *J. Med. Chem.*, **64** (2021), 8208–8220. https://doi.org/10.1021/acs.jmedchem.1c00020

17. H. Li, K. Sze, G. Lu, P. J. Ballester, Machine-learning scoring functions for structure-based drug lead optimization, *WIREs Comput. Mol. Sci.*, **10** (2020), e1465. https://doi.org/10.1002/wcms.1465

18. F. Miljkovic, R. Rodriguez-Perez, J. Bajorath, Machine learning models for accurate prediction of kinase inhibitors with different binding modes, *J. Med. Chem.*, **63** (2019), 8738–8748. https://doi.org/10.1021/acs.jmedchem.9b00867

19. Y. Lo, S. Rensi, W. Torng, R. B. Altman, Machine learning in chemoinformatics and drug discovery, *Drug Discovery Today*, **23** (2018), 1538–1546. https://doi.org/10.1016/j.drudis.2018.05.010

20. A. Abgeena, S. Garg, A novel convolution bi-directional gated recurrent unit neural network for emotion recognition in multichannel electroencephalogram signals, *Technol. Health Care*, **31** (2023), 1215–1234. https://doi.org/10.3233/THC-220458

21. P. Gong, C. Zhang, M. Chen, Deep learning for toxicity and disease prediction, *Front. Genet.*, **11** (2020), 175. https://doi.org/10.3389/fgene.2020.00175

22. A. Lavecchia, Deep learning in drug discovery: Opportunities, challenges and future prospects, *Drug Discovery Today*, **24** (2019), 2017–2032. https://doi.org/10.1016/j.drudis.2019.07.006

23. R. Dey, F. Salem, Gate-variants of gated recurrent unit (GRU) neural networks, in *2017 IEEE 60th International Midwest Symposium on Circuits and Systems*, IEEE, (2017), 1597–1600. https://doi.org/10.1109/MWSCAS.2017.8053243

24. L. Breiman, Random Forests, *Mach. Learn.*, **45** (2001), 5–32. https://doi.org/10.1023/A:1010933404324

25. K. Hsu, H. Gupta, S. Sorooshian, Artificial neural network modeling of the rainfall-runoff process, *Water Resour. Res.*, **31** (1995), 2517–2530. https://doi.org/10.1029/95WR01955

26. J. Friedman, T. Hastie, R. Tibshirani, Additive logistic regression: A statistical view of boosting (with discussion and a rejoinder by the authors), *Ann. Statist.*, **28** (2000), 337–407. https://doi.org/10.1214/aos/1016218223

27. S. Kim, J. Chen, T. Cheng, A. Gindulyte, J. He, S. He, et al., PubChem 2019 update: Improved access to chemical data, *Nucleic Acids Res.*, **47** (2019), D1102–D1109. https://doi.org/10.1093/nar/gky1033

28. Z. Sessions, N. Sánchez-Cruz, F. Prieto-Martínez, V. M. Alves, H. P. Santos Jr., E. Muratov, et al., Recent progress on cheminformatics approaches to epigenetic drug discovery, *Drug Discovery Today*, **25** (2020), 2268–2276. https://doi.org/10.1016/j.drudis.2020.09.021

29. T. Chen, C. Guestrin, Xgboost: A scalable tree boosting system, in *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, Association for Computing Machinery, (2016), 785–794. https://doi.org/10.1145/2939672.2939785

30. K. Yang, K. Swanson, W. Jin, C. Coley, P. Elden, H. Gao, et al., Analyzing learned molecular representations for property prediction, *J. Chem. Inf. Model.*, **59** (2019), 3370–3388. https://doi.org/10.1021/acs.jcim.9b00237

31. J. Durant, B. Leland, D. Henry, J. G. Nourse, Reoptimization of MDL keys for use in drug discovery, *J. Chem. Inf. Comput. Sci.*, **42** (2002), 1273–1280. https://doi.org/10.1021/ci010132r

32. S. Bagheri, M. Rahban, F. Bostanian, F. Esmaeilzadeh, A. Bagherabadi, S. Zolghadri, et al., Targeting protein kinases and epigenetic control as combinatorial therapy options for advanced prostate cancer treatment, *Pharmaceutics*, **14** (2022), 515. https://doi.org/10.3390/pharmaceutics14030515

AIMS Press