ISSN 2616-5775 Vol. 9, Issue 3: 72-81, DOI: 10.25236/AJCIS.2026.090309

Construction of a Graph Neural Network-Based Activity Prediction Model for Traditional Chinese Medicine Monomers Against Pancreatic Cancer

Qihao Wang[1], Yanxiang Xie[1], Bin Song[2,*]

[1]School of Health Science and Engineering, University of Shanghai for Science and Technology, Shanghai, 200093, China
[2]Department of Hepatobiliary, Pancreatic, and Spleen Surgery, Changhai Hospital, Naval Medical University, Shanghai, 200433, China
**Corresponding Author*

Abstract: *This study aims to screen Chinese herbal medicine monomers with anti-pancreatic cancer activity and establish a compound screening model using graph neural network deep learning algorithms to provide a novel strategy for large-scale screening of anti-pancreatic cancer herbal monomers, first employing CTG technology to detect the proliferation-inhibitory effects of 970 TCM monomers on BxPC-3 pancreatic cancer cells, then taking this data as training material to construct an anti-pancreatic cancer activity prediction model within the Chemprop framework, applying this model to predict anti-pancreatic cancer activity for over 30,000 natural product molecules in the TCMBank database, and conducting experimental validation on the top five TCM monomers with the highest predicted scores; the results showed that 87 of the 970 TCM monomers (approximately 9.0% of the total) exhibited >80% inhibition of BxPC-3 cell proliferation, the constructed model achieved an* R^2 *of 0.81 and an RMSE of 11.34 on the test set with excellent performance, and three of the five candidate compounds (Alkannin, Rottlerin, and* β*-Mangostin) selected for experimental validation exhibited significant, dose-dependent inhibitory effects on BxPC-3 cells; this study evaluated the anti-pancreatic cancer activity of 970 TCM monomers and used the results as a training set to construct a deep learning model, which enabled large-scale screening of TCM monomers for anti-pancreatic cancer activity and provided a novel strategy for AI-driven drug discovery.*

Keywords: *Traditional Chinese medicine; Pancreatic ductal adenocarcinoma; Graph neural networks; Deep learning; Anti-Cancer drug screens*

1. Introduction

Pancreatic ductal adenocarcinoma (PDAC) is one of the most aggressive and lethal malignancies of the digestive system, with a five-year survival rate of less than 10%[1]. Although recent advances in surgical resection, chemotherapy, and targeted therapy have led to modest improvements in patient outcomes, the overall efficacy of current treatment strategies remains limited due to the occult onset, strong drug resistance, and pronounced tumor heterogeneity of PDAC[2]. Consequently, the discovery of novel anti-pancreatic cancer agents has become an urgent clinical priority.

Natural products, particularly traditional Chinese medicine monomers (TCMMs), have long served as a valuable source of anticancer compounds owing to their structural diversity and wide range of biological activities[3]. However, conventional drug discovery approaches largely rely on extensive in vitro and in vivo experiments, which are time-consuming, labor-intensive, and inefficient when applied to large-scale compound libraries [4].

In recent years, artificial intelligence (AI), especially deep learning–based approaches, has demonstrated remarkable potential in the field of drug discovery. By modeling the relationships between large-scale molecular features and biological activity data, these methods can predict the potential activity of compounds during virtual screening, thereby significantly reducing the experimental workload and accelerating the identification of bioactive candidates [5,6]. Heid et al. developed the Chemprop model based on message passing neural networks (MPNNs). Taking molecular graphs as direct input, this model automatically learns structure–activity relationships through message passing between atoms and chemical bonds, and exhibits excellent prediction

ISSN 2616-5775 Vol. 9, Issue 3: 72-81, DOI: 10.25236/AJCIS.2026.090309

accuracy and generalization ability in tasks including molecular property prediction, drug screening, and toxicity assessment.Nevertheless, studies focusing on the prediction of anti-PDAC activity of natural products—particularly TCMMs—remain limited, and systematic screening frameworks that integrate computational prediction with experimental validation are still lacking.

In this study, we employed the Chemprop model to predict the anti-pancreatic cancer activity of traditional Chinese medicine monomers.The reliability of the model's predictions was further confirmed through in vitro experiments, leading to the preliminary identification of multiple TCMMs with potential anti-pancreatic cancer activity. This work provides a novel technical framework and effective tool for elucidating the pharmacological mechanisms of natural products and facilitating the discovery of new anti-pancreatic cancer therapeutics.

2. Materials and methods

2.1 Data Source of TCMMs and Evaluation of Anti-PDAC Activity

The TCMM compound library (Catalog No. L6810, June 2020 edition) was purchased from TOPSCIENCE. Detailed information for 970 compounds in the library, including their sources, structural types, molecular structures, targets, and signaling pathways, was collected, curated, and manually verified to establish a foundational database.

To evaluate the inhibitory effects of TCMMs on pancreatic cancer cells, the CellTiter-Glo (CTG) assay was employed to measure cell viability inhibition against BxPC-3 pancreatic cancer cells. Briefly, BxPC-3 cells in the logarithmic growth phase were harvested, counted, and adjusted to a concentration of 4.9×10^4 cells/mL. The suspension was serially diluted 1.25-fold to generate eight concentration gradients and seeded into 384-well plates (50 μL per well). Cells were incubated for 96 h at 37 ℃ in 5% CO_2. After incubation, 50 μL of CTG reagent was added to each well, followed by a 10-min incubation at room temperature, and luminescence was measured using a SPAPK microplate reader to determine the optimal seeding density.

For the formal experiment, the cell concentration was adjusted to 2.56×10^4 cells/mL, and 25 μL of cell suspension (approximately 640 cells/well) was seeded into 384-well plates and incubated overnight at 37°C in 5% CO_2. Subsequently, 25 μL of compound working solution was added to each well, and the plates were incubated for an additional 72 h. At the end of incubation, 50 μL of CTG reagent was added to each well, followed by 10 min at room temperature, and luminescence intensity was recorded using the SPAPK microplate reader. The inhibition rate (IR) was calculated according to the following formula:

$$IR(\%)=[1-\frac{RLU_S-RLU_{BLK}}{RLU_{NC}-RLU_{BLK}}]\times100\% \quad (1)$$

RLUS represents the luminance value of the sample well (cells + test compound); RLUNC denotes the luminance value of the negative control well (cells + solvent); RLUBLK indicates the luminance value of the blank well (culture medium + solvent).

2.2 Construction of the Chemprop Model

(1) Data Preparation and Dataset Partitioning

The training data for the model were derived from 970 samples in the traditional Chinese medicine monomer (TCMM) compound library, including their corresponding SMILES representations [7] and inhibitory activity data against pancreatic cancer cells. The dataset was randomly divided into a training set (80%), a validation set (10%), and a test set (10%) to ensure the fairness and stability of model training and evaluation.

(2) Model Construction and Training

In this study, a molecular property prediction model was constructed using the Chemprop framework [8]. Chemprop is a message passing neural network (MPNN)–based platform that represents each molecule as a graph consisting of atoms and chemical bonds. By simulating the message-passing process between atoms, the model automatically learns structure–activity relationships without the need for manually designed molecular fingerprints or descriptors. In recent years, Chemprop has been widely applied in drug screening, toxicity prediction, and molecular property evaluation, demonstrating superior prediction accuracy and generalization ability compared with traditional machine learning

ISSN 2616-5775 Vol. 9, Issue 3: 72-81, DOI: 10.25236/AJCIS.2026.090309

approaches [9–11]. The source code was obtained from the official GitHub repository (https://github.com/chemprop/chemprop, version 2.0.0), and model parameters were fine-tuned according to the training performance in this study.

During the training phase, SMILES strings and inhibition rates from the original dataset were first converted into the MoleculeDatapoint format. Subsequently, the SimpleMoleculeMolGraphFeaturizer module provided by Chemprop was used to generate graph-based molecular representations, followed by feature normalization to ensure consistent data distribution between the training and validation sets.

In terms of architecture, the model employed a bond-based message passing mechanism (BondMessagePassing) to encode molecular graphs, followed by mean aggregation (MeanAggregation) to integrate node-level features. The aggregated features were then passed through a regression-type feed-forward neural network (RegressionFFN) to produce the final prediction outputs. The mean squared error (MSE) was used as the loss function, while the root mean square error (RMSE) was adopted as the evaluation metric. An early stopping criterion based on validation RMSE was applied to prevent overfitting and ensure optimal generalization. Training was conducted for 100 epochs, with the loss and RMSE values on the validation set recorded after each epoch to monitor model convergence.

2.3 Model Prediction

(1) Source of Prediction Data

Prediction data were sourced from TCMBank [12], the world's largest repository of traditional Chinese medicine (TCM) ingredients. The full All-Ingredient dataset was downloaded from the TCMBank official website's Download section, containing approximately 61,966 TCM ingredient entries. Each entry included metadata such as TCMBank_ID, TCM_name, Molecular_Formula, SMILES, Molecular_Weight, Molecular_Volume, Ingredient_id, OB_score, CAS_id, SymMap_id, TCMID_id, TCMSP_id, TCM-ID_id, and PubChem_id—encompassing compound identifiers, source herbs, molecular formulas, SMILES strings, and key physicochemical properties.

(2) Data Preprocessing

Prior to model prediction, all SMILES strings from TCMBank were validated using RDKit (v2025.3.1). Compounds with invalid, unrecognized, or structurally inconsistent SMILES that violated fundamental organic chemical composition rules were removed. After preprocessing, approximately 30,000 structurally valid molecules were retained for subsequent activity prediction and analysis.

2.4 In Vitro Experiments

The in vitro assays were conducted using the Cell Counting Kit-8 (CCK-8) method to determine the IC_{50} values of five candidate compounds against BxPC-3 pancreatic cancer cells. Each compound was tested at six concentration levels: 50, 16.6667, 5.5556, 1.8519, 0.6173, and 0.2058 μM, with three replicate wells per concentration. The positive control compound, staurosporine (STSP), was tested at a concentration of 10 μM. Wells containing culture medium without cells served as blank controls, while wells containing only solvent-treated cells were used as negative controls, both with two replicates per group. All experiments were performed in 96-well plates.

(1) Cell Culture

BxPC-3 cells were passaged when cell confluence reached 80–90%. During subculturing, the old medium was removed, and the cells were rinsed with PBS, followed by digestion with trypsin. The digestion was terminated by adding a fivefold volume of complete culture medium. The cell suspension was collected and centrifuged at 200g for 5 minutes. After removing the supernatant, the cells were resuspended in 1 mL of fresh culture medium and subcultured at an appropriate ratio into new culture dishes. The cells were maintained at 37 ℃ in a humidified incubator with 5% CO_2.

(2) Cell Seeding and Compound Treatment

Cells in the logarithmic growth phase were collected and seeded into 96-well plates at a density of 6,000 cells per well in 50 μL of culture medium. The plates were incubated overnight at 37 ℃ with 5% CO_2 to allow cell attachment. After adhesion, 50 μL of compound working solution was added to each well, and the cells were incubated for an additional 72 hours. Following incubation, 10 μL of CCK-8 reagent was added to each well and incubated for 2 hours. The absorbance at 450 nm was then measured using a microplate reader to determine cell viability.

ISSN 2616-5775 Vol. 9, Issue 3: 72-81, DOI: 10.25236/AJCIS.2026.090309

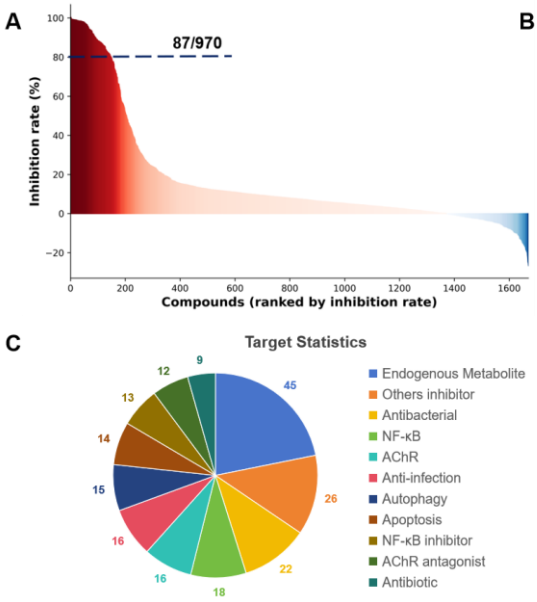
3. Results

3.1 Organization of the traditional Chinese medicine monomer library and screening for anti-pancreatic cancer activity

In this study, a traditional Chinese medicine monomer (TCMM) library (L6810) was purchased from TopScience. The library contains 970 monomeric compounds derived from more than 500 traditional Chinese medicines, covering various structural types of natural products, including alkaloids, flavonoids, terpenoids, and glycosides.

To evaluate the potential inhibitory effects of these compounds on pancreatic cancer cells, the CellTiter-Glo (CTG) assay was employed to screen all TCMMs for proliferation inhibition activity in a 384-well plate format. Human pancreatic cancer cells BxPC-3 were used as the experimental model, with all compounds tested at a final concentration of 10 μM. Staurosporine[13], a known cytotoxic agent, served as the positive control. The results showed that 87 TCMMs (approximately 9.0% of all compounds) exhibited more than 80% inhibition of BxPC-3 cell proliferation (Figure 1A). This indicates that certain TCMMs possess significant anti-pancreatic cancer potential and warrant further investigation.

In addition, the known signaling pathways and molecular targets of these TCMMs were organized and analyzed. The results revealed that, among the top 15 enriched pathways, those related to metabolism, apoptosis, and immune inflammation ranked the highest in compound counts (Figure 1B). The main targets included endogenous metabolites, antibacterial agents, and NF-κB (Figure 1C). These findings demonstrate the diversity and representativeness of the TCMM library, providing a solid foundation for subsequent systematic activity screening and mechanistic studies.



B

Pathways	Number
Metabolism	115
Microbiology&Virology	80
Immunology/Inflammation	74
oxidation-reduction	55
Apoptosis	52
Neuroscience	49
Membrane transporter/Ion channel	38
NF-Kb	33
DNA Damage/DNA Repair	26
Autophagy	19
GPCR/G Protein	17
Proteases/Proteasome	17
Cytoskeletal Signaling	15
PI3K/Akt/mTOR signaling	13
Endocrinology/Hormones	11

A: Ranking chart of the proliferation inhibition rates of 970 traditional Chinese medicine monomers on BxPC-3 pancreatic cancer cells. The horizontal axis represents the compounds sorted from lowest to highest inhibition rate, and the vertical axis represents the inhibition rate (%). B: Statistical table of Chinese herbal medicine single compounds targeting pathways. The first column represents the pathway category, and the second column represents the number of compounds targeting that pathway. C: Statistical pie chart of target types for traditional Chinese medicine monomers.

Figure 1. Distribution of anti-pancreatic cancer activity of traditional Chinese medicine monomers and target pathway information

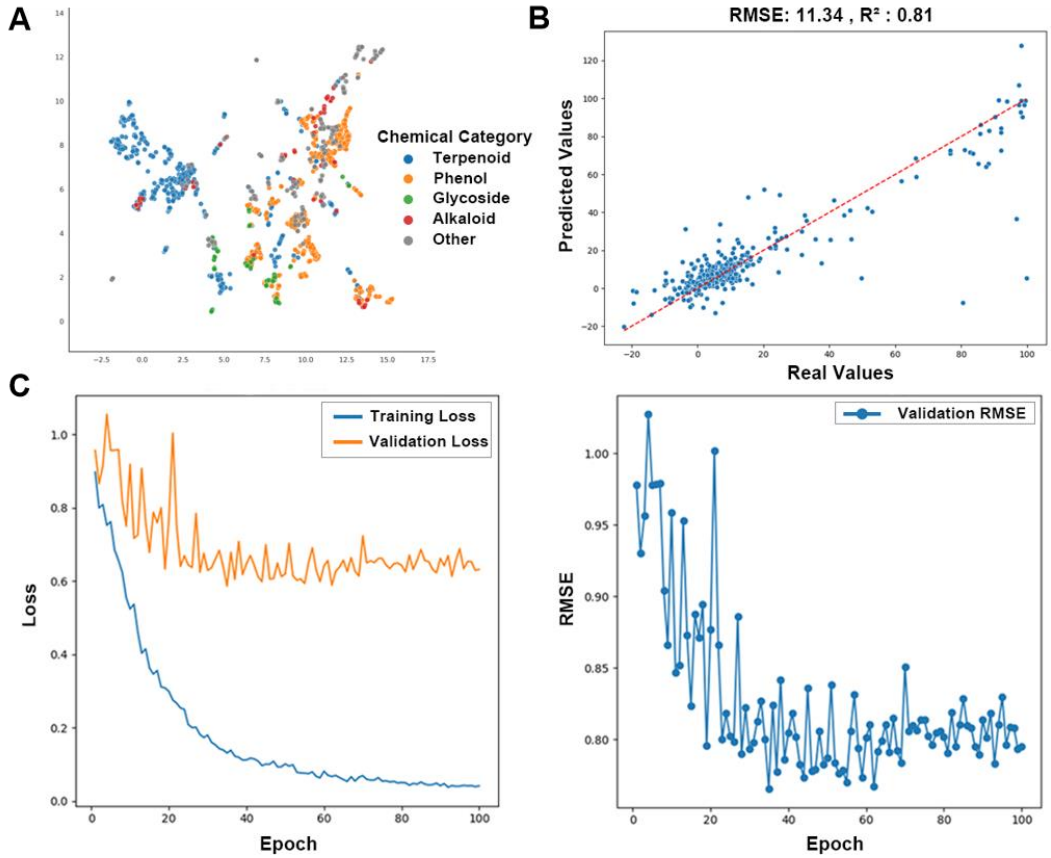
3.2 Construction and training of the Chemprop-based graph neural network model

For graph-structured deep learning methods, especially graph neural networks (GNNs) that directly take molecular graphs as input, the model performance largely depends on the structural diversity of the training data[14]. Therefore, before constructing the predictive model, it is necessary to systematically evaluate the structural diversity of the compound library used.

ISSN 2616-5775 Vol. 9, Issue 3: 72-81, DOI: 10.25236/AJCIS.2026.090309

To this end, we performed structural classification and feature dimensionality reduction on 970 traditional Chinese medicine monomers (TCMMs) to intuitively reveal their structural distribution characteristics and diversity. First, based on their SMILES representations and known SMARTS structural patterns[15], the molecular scaffolds were preliminarily classified into five major structural categories: terpenoids (354), phenolics (188), glycosides (113), alkaloids (75), and others that could not be assigned to the above categories (240). On this basis, ECFP4 molecular fingerprints[16] were calculated for all compounds, and the UMAP method[17] was applied for dimensionality reduction and visualization. The clustering plot (Figure 2A) showed that compounds of different structural types exhibited relatively distinct cluster distributions in two-dimensional space, with terpenoids and phenolics forming two major clusters, while glycosides and alkaloids showed partial overlap and transition. Overall, the compound library demonstrated reasonable coverage and differentiation in structural space, providing a solid data foundation for feature learning in subsequent modeling.

Next, we used the SMILES representations and pancreatic cancer cell inhibition rates of the 970 TCMMs as training data to construct a molecular graph–based regression model for predicting the potential inhibitory effects of new compounds on pancreatic cancer cells. To fully exploit molecular structural information and improve model generalization, the Chemprop framework was employed, and the dataset was randomly divided into training, validation, and test sets. The root mean square error (RMSE) and coefficient of determination (R^2) were used as the main evaluation metrics to assess prediction accuracy and model fitness.



A: UMAP dimensionality reduction visualization diagram of the structural diversity of TCM. B: Scatter plot of actual values and predicted values of the model on the test set. C: Loss curve of the model training set and validation set (left) and RMSE change curve of the validation set (right).

Figure 2. Distribution of training data structure types and model training performance

Evaluation on the independent test set showed a high consistency between predicted and experimental values, with an R^2 of 0.81 and an RMSE of 11.34 (Figure 2B). During training, the validation loss curve gradually decreased and stabilized, and the RMSE continuously declined (Figure 2C), indicating stable model convergence and low validation error. Collectively, these results demonstrate that the GNN-based deep learning model can accurately capture the structure–activity relationships of natural products, providing a reliable approach for screening potential anti-pancreatic cancer TCMM candidates.

ISSN 2616-5775 Vol. 9, Issue 3: 72-81, DOI: 10.25236/AJCIS.2026.090309

3.3 Model prediction and candidate compound screening

To verify the reliability and predictive capability of the constructed model, we introduced the knowledge graph platform iKraph[18] for external validation. iKraph is a large-scale biomedical knowledge graph designed for artificial intelligence–driven and data-driven biomedical research, capable of establishing multidimensional associations among diseases, drugs, targets, and molecular mechanisms.

First, using this platform, we screened a set of drugs known to be closely related to pancreatic cancer therapy, including 200 positive control molecules such as the first-line clinical drugs Gemcitabine and Irinotecan. Meanwhile, to ensure the objectivity of model discrimination, 200 drugs unrelated to pancreatic cancer treatment were randomly selected as negative controls.

After inputting the SMILES representations of the above compounds into the trained model for prediction, the results showed that the predicted scores of positive drugs were significantly higher than those of negative drugs, while the latter generally exhibited much lower values (Figure 3). This finding indicates that the model not only provides predictive scoring for new molecules but also possesses a strong ability to distinguish pancreatic cancer–related drugs from unrelated compounds, thereby further validating its effectiveness and application potential.

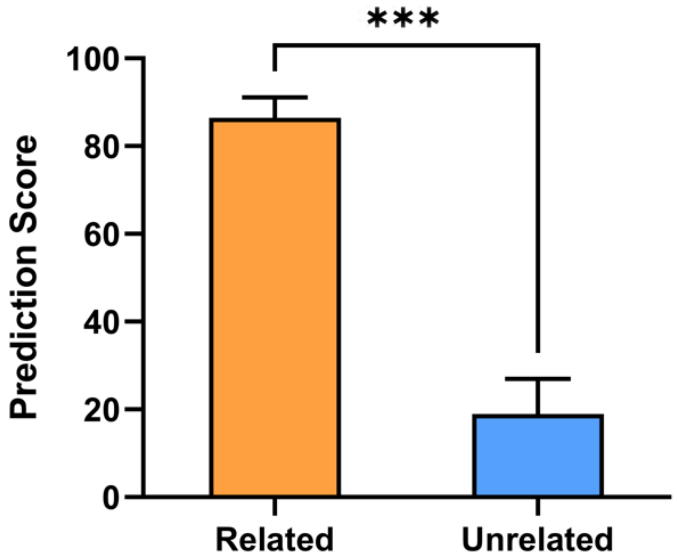


Figure 3. Prediction results for pancreatic cancer-related and non-related drugs

After constructing a well-performing regression model, we further applied it to a larger-scale library of traditional Chinese medicine monomers (TCMMs) to identify potential candidates with anti-pancreatic cancer activity. The compound library was obtained from TCMBank, the world's largest database of traditional Chinese medicine, which contains over 60,000 natural products. The SMILES representations of these compounds were filtered for validity and subsequently input into the trained model to predict the inhibitory activity scores against pancreatic cancer cells for each molecule.

From the high-scoring compounds with predicted values above 90, we further selected candidate molecules that have not yet been reported in pancreatic cancer–related studies but have been previously documented to possess antitumor, anti-inflammatory, or apoptosis-inducing activities. These were considered potential novel anti-pancreatic cancer TCMMs. Finally, five compounds were identified for subsequent in vitro experimental validation (Table 1).

Table 1. Information related to candidate traditional Chinese medicine monomers validated in vitro

Compound	Category	CAS ID	Known Activity
Alkannin	Naphthoquinones	517-88-4	Anti-inflammatory
β-Mangostin	Flavonoids	20931-37-7	Anti-cancer
Usnic acid	Diterpenoids	125-46-2	Cytotoxicity
Triptonide	Triterpenoid lactones	38647-11-9	Anti-cancer
Rottlerin	Polyphenols	82-08-6	Induction of apoptosis

ISSN 2616-5775 Vol. 9, Issue 3: 72-81, DOI: 10.25236/AJCIS.2026.090309

3.4 In vitro validation of anti-pancreatic cancer activity of candidate TCMMs

To verify the actual anti-pancreatic cancer activity of the candidate traditional Chinese medicine monomers (TCMMs) identified by the model, we conducted in vitro cell proliferation inhibition assays using the CCK-8 method[19] on the human pancreatic cancer cell line BxPC-3. A total of five candidate compounds—Alkannin, β-Mangostin, Usnic acid, Triptonide, and Rottlerin were tested. Six concentration gradients (50, 16.67, 5.56, 1.85, 0.62, and 0.21 μM) were applied for each compound, with three replicates per group, while the positive control Staurosporine (STSP) was set at 10 μM. All experiments were performed in 96-well plates, and after 72 hours of incubation, the absorbance at 450 nm was measured using a microplate reader to calculate the cell inhibition rate. The inhibitory effects of different compounds showed significant variation (Table 2, Fig. 4).

Table 2. In vitro inhibitory activity, IC50, and goodness of fit of candidate traditional Chinese medicine monomers

Compound	Max Inhibition Rate	**IC_{50} (μM)**	$LogIC_{50}$	R^2
β-Mangostin	99.34%	55.69	1.746	0.862
Alkannin	91.28%	0.105	-0.98	0.904
Rottlerin	86.46%	3.244	0.511	0.973
Triptonide	67.32%	143184	5.156	0.802
Usnic acid	42.74%	204535	5.311	0.782

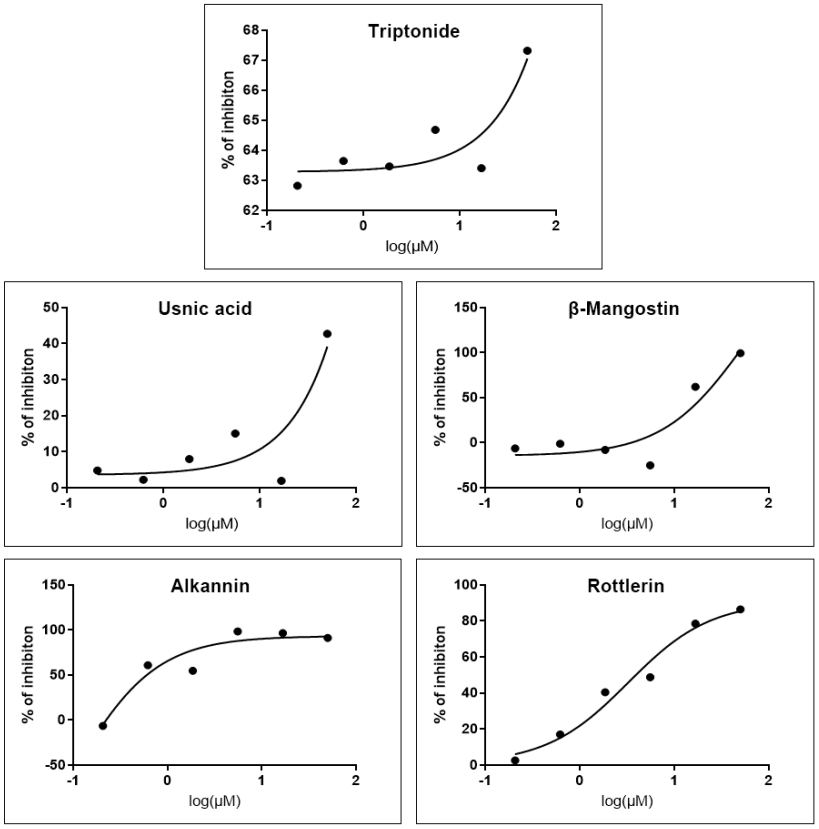


Figure 4. Dose-response fitting curves of five candidate traditional Chinese medicine monomers on BxPC-3 pancreatic cancer cells.

Triptonide exhibited a relatively constant inhibition rate of approximately 63%–68% across all tested concentrations, showing no evident dose dependence. Its fitted IC_{50} value reached 143,184 μM with $R^2 = 0.802$, indicating non-specific and weak inhibitory activity. Usnic acid also displayed limited inhibition, with a maximum inhibition rate of 42.74% and a fitted IC_{50} of 204,535 μM ($R^2 = 0.782$), suggesting a lack of significant antiproliferative activity under the tested conditions.

In contrast, Alkannin demonstrated a strong and dose-dependent inhibitory effect, with a maximum inhibition rate of 91.28%, a fitted IC_{50} of 0.10 μM, and $R^2 = 0.904$, indicating potent and stable cytotoxic activity. Rottlerin also showed a pronounced inhibitory trend, achieving a maximum

ISSN 2616-5775 Vol. 9, Issue 3: 72-81, DOI: 10.25236/AJCIS.2026.090309

inhibition rate of 86.46%, an IC_{50} of 3.244 μM, and a well-fitted dose–response curve (R^2 = 0.973). β-Mangostin exhibited minimal inhibition at low concentrations but reached a cell inhibition rate of 99.34% at high concentrations, with a fitted IC_{50} of 55.69 μM (R^2 = 0.862), indicating that its inhibitory effect requires relatively high doses to manifest.

Overall, all five candidate traditional Chinese medicine monomers (TCMMs) exerted inhibitory effects on pancreatic cancer cells to varying degrees, and three of them displayed potent antiproliferative activity in vitro. Alkannin and Rottlerin exhibited low IC_{50} values, distinct dose dependence, and strong inhibitory potency, supporting their potential as lead compounds for further mechanistic and pharmacological research. β-Mangostin showed significant growth inhibition at high concentrations but limited efficacy at low doses, warranting further exploration of its mechanism of action and dose-response relationship. In contrast, Triptonide and Usnic acid showed only moderate inhibitory activity under the tested conditions and are thus less promising for follow-up studies.

These experimental results not only validate the reliability of the graph neural network–based model for predicting the bioactivity of TCMMs but also furnish solid experimental evidence for identifying promising anti-pancreatic cancer drug candidates.

4. Discussion

In this study, a regression model was constructed based on Chemprop to predict the anti-pancreatic cancer activity of traditional Chinese medicine monomers (TCMMs). The model achieved high predictive accuracy on the test set (R^2 = 0.81, RMSE = 11.34), confirming the reliability and applicability of this approach in modeling structure–activity relationships of natural products. Compared with traditional machine learning methods that rely on manually crafted molecular descriptors, Chemprop directly takes molecular graphs as input and automatically extracts structure–activity features through message passing between atoms and bonds, thereby avoiding subjective bias in feature engineering. This advantage is particularly pronounced for TCMMs, which exhibit high structural diversity and chemical complexity, enabling the model to generalize well to unseen molecular structures and laying the foundation for discovering novel bioactive compounds.

At the prediction stage, the trained model was applied to the TCMBank database containing over 60,000 natural products. After structural validation and score filtering, a total of 143 candidate compounds with predicted activity scores above 90 were obtained. Literature mining revealed that several of these compounds have been previously reported to possess anti-pancreatic cancer activity, further validating the model's reliability. For example, Phenethyl isothiocyanate (PEITC) (predicted score = 98.23) has demonstrated significant antitumor efficacy both in vitro and in MIA PaCa-2 xenograft mouse models by inducing G2/M-phase arrest, regulating apoptotic proteins and inhibiting the Notch1/2 pathway. Similarly, α-Mangostin (predicted score = 95.16) effectively inhibits pancreatic cancer cell proliferation by inducing apoptosis, suppressing PI3K/Akt and NF-κB signaling and inhibiting epithelial-mesenchymal transition (EMT). In addition, 10-Hydroxycamptothecine and its prodrug irinotecan also exhibit strong cytotoxicity against pancreatic cancer cells, which is consistent with the model's high prediction scores.

From the prediction results, five TCMMs without previous pancreatic cancer studies but with known antitumor or related bioactivities were selected for in vitro validation. The CCK-8 assay revealed that three compounds—Alkannin, Rottlerin, and β-Mangostin—exhibited reliable inhibitory effects on BxPC-3 cells. Among them, Alkannin and Rottlerin showed pronounced and consistent inhibition across multiple concentration gradients, with low IC50 values and high goodness-of-fit values (R2> 0.90), indicating a clear dose-dependent response, which suggests that these compounds can exert strong antiproliferative effects even at low concentrations and have potential for further development. In contrast, β-Mangostin nearly completely suppressed cell proliferation at high concentrations but showed limited activity at low doses, with a relatively high IC50 (55.69 μM). This may be attributed to its limited cellular uptake, weak target affinity, or low metabolic stability, and its clinical potential may depend on the optimization of dosage and delivery strategies. Triptonide and Usnic acid showed only moderate inhibitory activity under the tested conditions and are thus less promising for follow-up studies.

Despite these encouraging results, certain limitations remain. Experimental validation in this study primarily relied on in vitro cellular models, whereas pancreatic cancer exhibits high complexity and heterogeneity in clinical settings. In future work, both in vitro and in vivo studies will be emphasized,

ISSN 2616-5775 Vol. 9, Issue 3: 72-81, DOI: 10.25236/AJCIS.2026.090309

including evaluations in mouse xenograft and pancreatic organoid models to further assess the antitumor efficacy of candidate TCMMs. Experimental results will be iteratively integrated into the model to refine prediction performance, while multi-omics approaches such as transcriptomics and metabolomics will be employed to elucidate molecular mechanisms of action.

5. Conclusion

This study successfully established a Chemprop-based graph neural network regression model for the prediction of anti-pancreatic cancer activity of traditional Chinese medicine monomers (TCMMs). The model exhibited excellent predictive performance with an R^2 of 0.81 and an RMSE of 11.34 on the independent test set, and its strong discrimination ability was further verified by external validation with pancreatic cancer-related and unrelated drugs. By applying this model to the TCMBank database, we completed anti-pancreatic cancer activity prediction for more than 30,000 structurally valid natural product molecules, and screened five candidate compounds for in vitro experimental validation. Among them, Alkannin, Rottlerin and β-Mangostin were identified as potential anti-pancreatic cancer active components, with Alkannin showing the most potent activity (IC50=0.105 μM) and a significant dose-dependent inhibitory effect on BxPC-3 cells, which is a promising lead compound for subsequent anti-pancreatic cancer drug development.

This study innovatively combines graph neural network technology with traditional Chinese medicine monomer screening, which significantly improves the efficiency and accuracy of identifying anti-pancreatic cancer active components from large-scale natural product libraries compared with traditional experimental screening methods. The constructed model provides a reliable technical tool for AI-driven anti-pancreatic cancer drug discovery, and also verifies the application value of deep learning in the modernization research of traditional Chinese medicine, providing new ideas for the excavation and development of natural active products.

For the follow-up research, the in vivo antitumor efficacy and action mechanisms of the screened active TCMMs will be further explored through animal models and multi-omics technologies. Meanwhile, the prediction model will be continuously optimized by integrating more experimental data and molecular information to further improve its generalization ability and prediction accuracy. It is expected to promote the transformation of these potential active components into preclinical and clinical research, and provide new therapeutic options for the clinical treatment of pancreatic cancer with high malignancy and poor prognosis.

References

[1] SIEGEL R L, GIAQUINTO A N, JEMAL A. Cancer statistics, 2024[J/OL]. CA: A Cancer Journal for Clinicians, 2024, 74(1): 12-49. DOI:10.3322/caac.21820.

[2] CONNOR A A, GALLINGER S. Pancreatic cancer evolution and heterogeneity: integrating omics and clinical data[J/OL]. Nature Reviews Cancer, 2022, 22(3): 131-142. DOI:10.1038/s41568-021-00418-1.

[3] ZHAO W, ZHENG X D, TANG P Y Z, et al. Advances of antitumor drug discovery in traditional Chinese medicine and natural active products by using multi-active components combination [J/OL]. [2025] [2025-08-25]. https://onlinelibrary.wiley.com/doi/10.1002/med.21963.DOI:10.1002/med.21963.

[4] AN Q, HUANG L, WANG C, et al. New strategies to enhance the efficiency and precision of drug discovery [J/OL]. Frontiers in Pharmacology, 2025, 16[2025-08-25]. https://www.frontiersin.org/journals/ pharmacology/articles/10.3389/fphar.2025.1550158/full. DOI:10.3389/fphar.2025.1550158.

[5] VAMATHEVAN J, CLARK D, CZODROWSKI P, et al. Applications of machine learning in drug discovery and development[J/OL]. Nature Reviews Drug Discovery, 2019, 18(6): 463-477. DOI:10.1038/s41573-019-0024-5.

[6] YANG K, SWANSON K, JIN W, et al. Analyzing Learned Molecular Representations for Property Prediction[J/OL]. Journal of Chemical Information and Modeling, 2019[2025-08-25]. https://pubs.acs.org/ doi/full/10.1021/acs.jcim.9b00237. DOI:10.1021/acs.jcim.9b00237.

[7] WEININGER D, WEININGER A. SMILES, a chemical language and information system. 1. Introduction to methodology and encoding rules [J]. Journal of Chemical Information and Computer Sciences, 1988, 28 (1): 31-36. DOI:10.1021/ci00057a005.

[8] HEID E, GREENMAN K P, CHUNG Y, et al. Chemprop: A Machine Learning Package for

ISSN 2616-5775 Vol. 9, Issue 3: 72-81, DOI: 10.25236/AJCIS.2026.090309

Chemical Property Prediction[J/OL]. Journal of Chemical Information and Modeling, 2023 [2025-08-25]. https://pubs.acs.org/doi/full/10.1021/acs.jcim.3c01250. DOI:10.1021/acs.jcim.3c01250.
[9] WONG F, OMORI S, DONGHIA N M, et al. Discovering small-molecule senolytics with deep neural networks[J/OL]. Nature Aging, 2023, 3(6): 734-750. DOI:10.1038/s43587-023-00415-z.
[10] NOTWELL J H, WOOD M W. ADMET property prediction through combinations of molecular fingerprints[A/OL]. arXiv, 2023[2025-08-25]. http://arxiv.org/abs/2310.00174. DOI:10.48550/arXiv. 2310.00174.
[11] ISERT C, KROMANN J C, STIEFL N, et al. Machine Learning for Fast, Quantum Mechanics-Based Approximation of Drug Lipophilicity[J/OL]. ACS Omega, 2023[2025-08-25]. https://pubs.acs.org/doi/full/ 10.1021/acsomega.2c05607. DOI:10.1021/acsomega.2c05607.
[12] LV Q, CHEN G, HE H, et al. TCMBank-the largest TCM database provides deep learning-based Chinese-Western medicine exclusion prediction[J/OL]. Signal Transduction and Targeted Therapy, 2023, 8(1): 127. DOI:10.1038/s41392-023-01339-1.
[13] OMURA S, SASAKI Y, IWAI Y, et al. Staurosporine, a Potentially Important Gift from a Microorganism[J/OL]. The Journal of Antibiotics, 1995, 48(7): 535-548. DOI:10.7164/antibiotics.48.535.
[14] ZHANG Y, SUI X, PAN F, et al. A comprehensive large-scale biomedical knowledge graph for AI-powered data-driven biomedical research[J/OL]. Nature Machine Intelligence, 2025, 7(4): 602-614. DOI:10.1038/s42256-025-01014-w.
[15] FANG J, LIU D, TSE C K. Impact of Structure of Network Based Data on Performance of Graph Neural Networks[C/OL]//2023 IEEE International Symposium on Circuits and Systems (ISCAS). 2023: 1-5 [2025-08-25]. https://ieeexplore.ieee.org/abstract/document/10182188. DOI:10.1109/ISCAS46773.2023.10182188.
[16] EDWARDS W, BARRON F H. SMARTS and SMARTER: Improved Simple Methods for Multiattribute Utility Measurement[J/OL]. Organizational Behavior and Human Decision Processes, 1994, 60(3): 306-325. DOI:10.1006/obhd.1994.1087.
[17] PROBST D, REYMOND J L. A probabilistic molecular fingerprint for big data settings[J/OL]. Journal of Cheminformatics, 2018, 10(1): 66. DOI:10.1186/s13321-018-0321-8.
[18] MCINNES L, HEALY J, MELVILLE J. UMAP: Uniform Manifold Approximation and Projection for Dimension Reduction[A/OL]. arXiv, 2020[2025-08-25]. http://arxiv.org/abs/1802.03426. DOI:10.48550/arXiv.1802.03426.
[19] CAI L, QIN X, XU Z, et al. Comparison of Cytotoxicity Evaluation of Anticancer Drugs between Real-Time Cell Analysis and CCK-8 Method[J/OL]. ACS Omega, 2019[2025-08-25]. https://pubs.acs.org/doi/full/10.1021/acsomega.9b01142. DOI:10.1021/acsomega.9b01142.