Synthetic and Systems Biotechnology 10 (2025) 148–155

Contents lists available at ScienceDirect

Synthetic and Systems Biotechnology

journal homepage: www.keaipublishing.com/en/journals/synthetic-and-systems-biotechnology

Original Research Article

Predicting effective drug combinations for cancer treatment using a graph-based approach

Qi Wang [a], Xiya Liu [d], Guiying Yan [b,c,*]

[a] College of Science, China Agricultural University, Beijing, 100083, China
[b] Academy of Mathematics and Systems Science, Chinese Academy of Sciences, Beijing, 100190, China
[c] University of Chinese Academy of Sciences, Beijing, 100190, China
[d] Institute of Biophysics, Chinese Academy of Sciences, Beijing, 100101, China

ARTICLE INFO

Keywords:
Drug combination
Cancer therapy
Computational method
Random walk with restart

ABSTRACT

Drug combination therapy, involving the use of two or more drugs, has been widely employed to treat complex diseases such as cancer. It enhances therapeutic efficacy, reduces drug resistance, and minimizes side effects. However, traditional methods to identify effective drug combinations are time-consuming, costly, and less efficient than computational methods. Therefore, developing computational approaches to predict drug combinations has become increasingly important.

In this paper, we developed the Random Walk with Restart for Drug Combination (RWRDC) model to predict effective drug combinations for cancer therapy. The RWRDC model offers a quantitative mathematical method for predicting the potential effective drug combinations. Cross-validation results indicate that the RWRDC model outperforms other predictive models, particularly in breast, colorectal, and lung cancer predictions across various performance metrics. We have theoretically proven the convergence of its algorithm and provided an explanation for the algorithm's rationality. A targeted case study on breast cancer further highlights the capability of RWRDC to identify effective drug combinations. These findings highlight our model as a novel and effective tool for discovering potential effective drug combinations, offering new possibilities in therapy. Additionally, the graph-based framework of RWRDC holds potential for predicting drug combinations in other complex diseases, expanding its utility in the medical field.

1. Introduction

Drug combinations, involving the integration of two or more active pharmaceutical ingredients into a single dosage form, have become increasingly vital in enhancing therapeutic efficacy while reducing the side effects and resistance associated with monotherapies [1]. They are classified as effective or ineffective based on demonstrated improvements in clinical trials or preclinical studies [2]. Notably, effective drug combinations interact with multiple targets within the molecular networks of complex diseases, making them pivotal for treating a wide array of conditions, including various cancers [3], HIV [4], and bacterial infections [5].

Given the vast number of potential drug combinations, experimental screening is both time-consuming and costly. This has underscored the need for computational methods in drug screening. Zhao et al. reviewed databases, web servers, and computational models for drug-drug interaction predictions, highlighting their potential to enhance the efficiency of drug combination screenings [6]. The rise of artificial intelligence has brought significant advancements to the field of drug combination prediction [7,8]. For example, Chen et al. developed the NLLSS model, which was validated through subsequent biological experiments, confirming 7 out of 13 predicted antifungal synergistic drug combinations [9]. Other studies, such as those by Hansen et al. [10] and Ali et al. [11], have employed algorithms like Random Forest and Support Vector Machines to identify effective drug pairs for cancer treatment. Huang et al. [12] used logistic regression models to predict effective drug combinations based on clinical phenotype information, while Xu et al. [13] constructed an SGB prediction model utilizing various features for predicting combinations of FDA-approved antihypertensive drugs. Han et al. [14] predicted drug-drug interactions by integrating deep learning

Peer review under responsibility of KeAi Communications Co., Ltd.
* Corresponding author. Academy of Mathematics and Systems Science, Chinese Academy of Sciences, Beijing, 100190, China.
E-mail address: yangy@amss.ac.cn (G. Yan).

https://doi.org/10.1016/j.synbio.2024.09.003
Received 26 June 2024; Received in revised form 17 August 2024; Accepted 8 September 2024
Available online 11 September 2024

with biomedical knowledge graphs.

In recent years, the significance of graph networks in bioinformatics research has been increasingly recognized. These networks play a crucial role in systematically exploring the molecular complexities associated with specific diseases. They enable the identification of disease modules and signaling pathways, facilitating the investigation of potential drug mechanisms [15]. Graph network methods, which focus on node connectivity, interactions, and relationships, have proven effective in addressing drug-disease interactions that are biologically relevant and clinically significant [16]. Their applications have extended to the study of drug combinations, where they have demonstrated versatility and potential in this evolving field.

Due to the limitations in obtaining labeled data, especially the scarcity of marked negative samples (i.e., known non-effective drug combinations), the effectiveness of traditional supervised learning methods is significantly impaired. To overcome this challenge, we have developed the Random Walk with Restart for Drug Combination (RWRDC) model, a novel graph-based semi-supervised learning framework, to predict potential effective drug combinations.

The efficacy of the RWRDC model is demonstrated by its performance in extensive cross-validation tests, consistently outperforming other methods, especially in the prediction of breast, colorectal, and lung cancers, across several performance metrics. We have theoretically proven the convergence of its algorithm and provided an explanation for the algorithm's rationality. A targeted case study on breast cancer exemplifies the RWRDC's capability to identify potential drug combinations. The paper concludes with an in-depth discussion on the strengths and limitations of the RWRDC model, as well as its potential applications in diverse medical scenarios.

2. Methods

2.1. Data collection and pre-processing

The Drug Combination Database (DCDB) includes information on 1363 unique drug combinations, involving 1735 individual drugs. Notably, the DCDB archives a total of 1103 effective combinations and 237 non-effective ones, offering a rich data source for identifying potential drug combinations. In our study, we classify effective drug combinations as positive samples and non-effective combinations as negative samples.

Due to the high mortality rates associated with breast, colorectal, and lung cancers, as referenced [17], we specifically focus on drug combinations related to these types of cancer as detailed in the DCDB. Our selection criteria are based on the use of ICD-10, a globally recognized disease classification system developed by the WHO [18]. Each drug combination in DCDB is tagged with one or more ICD-10 codes, facilitating targeted data extraction. For instance, combinations coded as C18 (malignant neoplasm of the colon) were earmarked as colorectal cancer data. Similarly, codes C50 and C34 were indicative of breast and lung cancer data, respectively. In this study, we excluded drug combinations from the DCDB database that are marked as 'Need further study'.

Here, the edge refers to an effective drug combination, and vertex denotes the drugs involved in these combinations. Density is defined as the ratio of the number of edges to the total number of possible edges, signifying the proportion of effective combinations among all feasible pairings. Table 1 shows that only 5 %–9 % of all possible drug pairings are potentially effective. Furthermore, the datasets for breast cancer, colorectal cancer, and lung cancer contain 4, 1, and 2 known non-effective combinations, respectively, which correspond to the number of drugs involved in effective combinations for each cancer type. These negative samples are utilized to validate the reasonableness of the effective scores discussed in *section B* of the Results.

Table 1
Basic information of the drug combination datasets.

Datasets	Number of vertices	Number of edges	Density
Breast cancer	41	46	0.0561
Colorectal cancer	18	13	0.0850
Lung cancer	24	22	0.0797

2.2. Notations and definitions

Let $G(V,E)$ be an undirected graph, where the vertex set V represents individual drugs, and the edge set E represents pairs of effective drug combinations, defined as $E = \{(v_i, v_j) \mid v_i \textit{ and } v_j \textit{ are a pair of effective drug combination}\}$.

Define the set of positive samples as $PS = \{(v_i,v_j)|(v_i,v_j) \in E\}$ and the set of negative samples as $NS = \{(v_i,v_j)| v_i \textit{ and } v_j \textit{ are a pair of non} - \textit{effective drug combination}\}$. Let the set of candidate samples be $CS = \{(v_i,v_j)| (v_i,v_j) \notin E\}$.

2.3. Random walk with restart algorithm

We are provided with a connected graph G where the transition probability p_{ij} represent the likelihood of moving from vertex j to i. In a standard Random Walk, at each stage, movement to any adjacent vertex is possible. The primary goal is to determine the most probable vertex we will end up at after a series of steps. We regard the vertex set $V = \{v_1, v_2, \ldots, v_n\}$ as a set of states $S = \{s_1, s_2, \ldots, s_n\}$ in a finite Markov chain $\mathscr{M}$, where each vertex corresponds to a state. For any v_j in V, we have $\sum_{v_i \in V} p_{ij} = 1$, allowing us to define a transition probability matrix $P = \{p_{ij}\} \in \mathbb{R}^{|V|\times|V|}$ of $\mathscr{M}$.

The random walk from v_j to v_i of graph G is to choose an edge e_{ij} randomly. Define transition probability p_{ij} as follows:

$$p_{ij} = \frac{1}{d_j}, \tag{1}$$

where d_j is the degree of vertex v_j. We denote D_G to be the diagonal matrix containing the vertex degrees of the graph and A to be the adjacency matrix of G. Thus, P can be rewritten in matrix notation as follows:

$$P = D_G^{-1}A. \tag{2}$$

Let $r_t \in \mathbb{R}^{|V|\times 1}$ be a vector in which the i-th element represents the probability of finding the random walk at vertex v_i at step t. The probability r_{t+1} can be calculated iteratively by:

$$r_{t+1} = P^T r_t. \tag{3}$$

In random walk with restart (RWR) the main difference is that at each stage there is a probability that we can restart i.e. go to the starting node [19]. For the RWR algorithm, there is an additional restart item compared to the above algorithm. The probability r_{t+1} can be calculated iteratively by the following expression:

$$r_{t+1} = cP^T r_t + (1-c)r_0. \tag{4}$$

Define the initial probability vector $r_0 \in \mathbb{R}^{|V|\times 1}$ such that the i-th element is 1 if it corresponds to an initial vertex, and 0 otherwise. Here, c is the probability of continuing the walk, making $1-c$ the restart probability, which ranges from 0 to 1 inclusive.

At the beginning of the RWR, we select a starting vertex v_i, and the walk proceeds based on transition probability. Suppose the walk proceeds arrive at vertex v_j; it then has a probability c of continuing based on the transition probability and a probability of $1-c$ of restarting the walk, that is, returning to vertex v_i. After several steps, the RWR reaches a stable state, when the $L1$ norm between r_{t+1} and r_t is less than an arbitrarily small value. Here, we set the cutoff at 10^{-6}, a criterion is

commonly used in various research studies [20,21]. When the RWR is stable, the stable probability between vertex i and vertex j is defined as the j-th element of r_t, given that the starting vertex is v_i. Fig. 1 illustrates the flowchart of the RWR process.

2.4. .Proof of convergence

This section demonstrates the convergence of the Random Walk with Restart (RWR) algorithm, demonstrating that as t approaches infinity, the steady-state condition $r_{t+1} = r_t$ is satisfied, according to equation (4). To clearly demonstrate this, let us define:

$$X = cP^T. \tag{5}$$

$$Y = (1-c)\left(I - cP^T\right)^{-1}. \tag{6}$$

Applying equations (5) and (6), we derive the following relationship:

$$r_{t+1} - Yr_0 = X(r_t - Yr_0). \tag{7}$$

By defining $M_t = r_t - Yr_0$, we can then express the next iteration as:

$$M_{t+1} = XM_t. \tag{8}$$

From Equation (8), setting $t = 0$ yields $B_0 = (I - Y)r_0$, which leads to:

$$M_t = X^t(I - Y)r_0, \tag{9}$$

$$r_t = [Y + X^t(I - Y)]r_0. \tag{10}$$

The limit of X^t as t approaches infinity is 0, i.e., $\lim_{t\to\infty} X^t = 0$, so:

$$\lim_{t\to\infty} r_t = Yr_0 = (1-c)\left(I - cP^T\right)^{-1} r_0. \tag{11}$$

Therefore, we conclude that:

$$\lim_{t\to\infty}(r_{t+1} - r_t) = 0. \tag{12}$$

This confirms the convergence of the RWR algorithm.

2.5. Effective score of drug-drug pairs

Each candidate drug-drug pair receives two rounds of effective scoring through the Random Walk with Restart (RWR) method. Define $\Pi = \{\pi_{ij}\} \in \mathbb{R}^{|V|\times|V|}$ be the stable probability matrix, where π_{ij} signifies the stable probability between vertices v_i and v_j. This means when RWR initiates from vertex v_i, π_{ij} represents the probability of reaching vertex v_j as the process stabilizes.

Specifically, π_{ij} is the j-th element of the steady-state probability ($\lim_{t\to\infty} r_t$) given that the i-th element of r_0 is 1. The value of π_{ij} represents the likelihood of an effective combination between drug j (among $n-1$ other candidate drugs excluding drug i) and the given drug i; likewise, π_{ji} indicates the potential for an effective combination between drug i (excluding drug j) and the given drug j. This study aims to predict effective drug pairings; hence each drug cannot be combined with itself but can potentially combine with any of the other $n-1$ drugs. We define the effective score $S = \{s_{ij}\} \in \mathbb{R}^{|V|\times|V|}$ as follows:

$$s_{ij} = \begin{cases} \pi_{ij} + \pi_{ji} & \text{if } i \neq j \\ 0 & \text{if } i = j \end{cases} \tag{13}$$

The Random Walk with Restart for Drug Combination (RWRDC) model can calculate a score for any drug pair. Pairs with higher scores are more likely to be effective drug combinations.

3. Results

Firstly, we analyzed the selection of restart probability values and the rationality of the scoring mechanism within the Random Walk with Restart for Drug Combination (RWRDC) model. Then, we analyzed the predictive performance of RWRDC through cross-validation. Finally, a case study verified the model's effectiveness in discovering novel drug combinations. The study highlighted RWRDC's practical utility and potential in facilitating the discovery of effective drug combination therapy.

3.1. Effects of restart probability

The receiver operating characteristic (ROC) curve is a graphical representation that illustrates the diagnostic capability of a binary classifier system across varying threshold levels. The Area Under the Receiver Operating Characteristic Curve (AUC) quantifies this capability, with values ranging from 0 to 1. A larger AUC value signifies a more effective classifier.

Leave-One-Out Cross-Validation (LOOCV), particularly advantageous for smaller datasets, involves sequentially using each positive sample (a known effective drug combination) as the test set, while the remaining positive samples and other non-positive samples (potential drug combinations) form the training set. The Random Walk with Restart for Drug Combination (RWRDC) model scores and ranks each sample in the positive samples and candidate samples. This process is repeated for each positive sample, with the model's predictive performance being evaluated against varying thresholds.

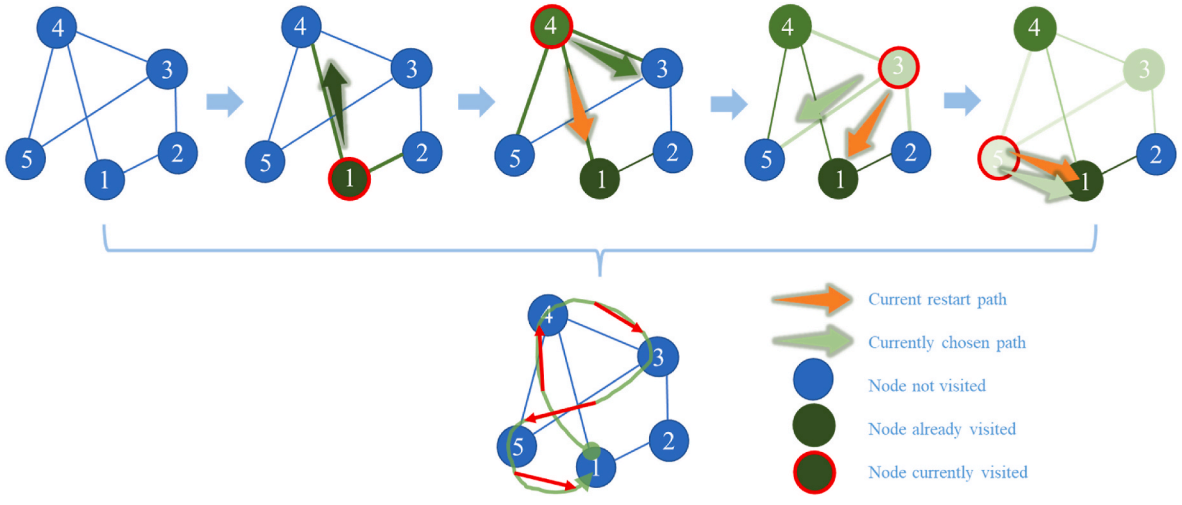

Fig. 1. Flowchart of the random walk with restart.

We investigated the impact of the restart probability $(1-c)$ on the RWRDC's performance by computing AUCs for values ranging from 0.05 to 0.95 using LOOCV. As depicted in Fig. 2, the model's efficacy, indicated by the average AUC (represented by the orange line), was maximized at a restart probability around 0.05, illustrating the importance of this parameter in enhancing model performance. Therefore, we set $(1-c)$ to be 0.05.

3.2. Rationality analysis of scoring

To check whether the drug-drug pairs with high effective scores were more likely to be effectively combined, all candidate drug-drug pairs in three datasets were ranked by RWRDC. In this paper, the average effective score of *PS* (effective combinations), *CS* (unknown combinations), and *NS* (non-effective combinations in candidate samples) were calculated respectively. If the average effective score of *PS* was higher than that of *CS* and also the average effective score of *CS* higher than that of *NS* on different datasets, the scoring process was reasonable.

The average effective scores of *PS*, *CS*, and *NS* were expressed by *PS score*, *CS score*, and *NS score*, respectively. As shown in Table 2, the PS scores were higher than the CS scores, and the CS scores were higher than the NS scores. Drug-drug pairs with higher effective scores were more likely to be effective combinations across the three datasets.

3.3. Cross-validation tests

In this paper, we used LOOCV for different methods and datasets to test the predictive ability of the model. Briefly, the drug-drug pairs corresponding to the elements of 1 in the adjacent matrix A were proportionally divided into the training set and the test set, and the elements of 0 in the adjacent matrix A were used as the *CS*. Then, the effective score of each sample of the test set and the *CS* was predicted.

Furthermore, our study involved a comparative analysis of the Random Walk with Restart for Drug Combinations (RWRDC) model against not only the conventional Random Walk (RW) model and the Network-based Laplacian regularized Least Square Synergistic drug combination prediction (NLLSS) model but also several widely-used machine learning algorithms, such as Logistic Regression (LR), Support Vector Machine (SVM), Random Forest (RF), and Stochastic Gradient Descent Classifier (SGDC) across various datasets. To ensure a fair comparison of the predictive ability of these models, we used the same known effective drug combination data as model inputs.

We employed LOOCV to ensure a fair performance evaluation in all datasets. The performance was evaluated using a comprehensive set of metrics, including accuracy, recall (sensitivity), specificity, precision, F-Score, MCC, AUC, and AUPR. In particular, the AUPR is a measure of the average precision of a classification model at different threshold settings. It is the area under the curve when the precision is plotted against

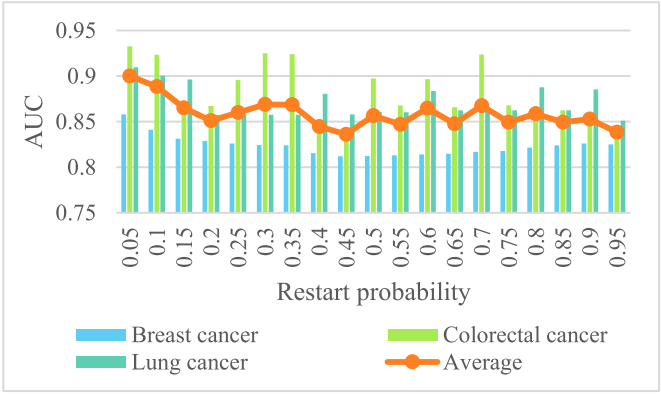

Fig. 2. The performance of restart probability on the LOOCV.

Table 2
Effective scores comparison across different datasets.

Datasets	*PS score*	*CS score*	*NS score*
Breast cancer	$4.25^{*}\ 10^{-2}$	$1.56^{*}\ 10^{-4}$	$6.96^{*}\ 10^{-5}$
Colorectal cancer	$6.59^{*}\ 10^{-2}$	$2.27^{*}\ 10^{-4}$	$5.96^{*}\ 10^{-4}$
Lung cancer	$5.19^{*}\ 10^{-2}$	$3.16^{*}\ 10^{-4}$	$1.82^{*}\ 10^{-7}$

the recall for various threshold levels. The precision-recall curve is especially useful in imbalanced classification problems. And other metrics we used are defined as:

$$Accuracy = \frac{TP + TN}{TP + TN + FN + FP} \tag{14}$$

$$Recall = \frac{TP}{TP + FN} \tag{15}$$

$$Specificity = \frac{TN}{TN + FP} \tag{16}$$

$$Precision = \frac{TP}{TP + FP} \tag{17}$$

$$F\ -\ Score = \frac{2 \times \text{Recall} \times \text{Precision}}{\text{Recall} + \text{Precision}} \tag{18}$$

$$MCC = \frac{TP \times TN\ -\ FN \times FP}{\sqrt{(TN + FN) \times (TP + FP) \times (TP + FN) \times (TN + FP)}} \tag{19}$$

where *TP* is the number of correctly predicted effective drug combinations, *TN* is the number of correctly predicted non-effective drug combinations, *FP* is the number of non-effective drug combinations predicted as effective ones, and *FN* is the number of effective drug combinations wrongly predicted as non-effective ones.

Table 3 shows a detailed evaluation of several classification models across different cancer datasets, using the LOOCV method to ensure the reliability and generalizability. Our RWRDC model, especially in lung and colorectal cancer datasets, demonstrates excellent performance in metrics such as Accuracy, Recall, Specificity, Precision, F-Score, and MCC. It notably outperforms other models, highlighting its robustness in predicting effective drug combinations. The AUC and AUPR values further confirm the superior predictive accuracy of RWRDC, especially in handling class imbalances.

The table includes *p*-values, providing statistical evidence of RWRDC's performance superiority over other methods, particularly noticeable in lung and colorectal cancers.

While the RWRDC model demonstrates strong performance across multiple metrics, it does not achieve the highest Accuracy and Specificity in the Breast and Colorectal Cancer Datasets. This suggests a need for further refinement in feature selection and model generalization to better adapt to the complexities of these datasets. In the Lung Cancer Dataset, the model's Recall is not optimal, indicating potential misses of actual positive cases. This may stem from the high variability inherent in lung cancer data, which calls for a more tailored approach in feature representation and model training to enhance accuracy.

3.4. Case study

This study entailed an exhaustive case analysis to evaluate the Random Walk with Restart for Drug Combination (RWRDC) model in uncovering novel, effective drug combinations, with a specific emphasis on breast cancer. The process ranked all possible drug pairings for targeted disease treatment based on their estimated effective scores from the RWRDC model. During the validation phase, we compared these scores against the Drug Combination Database (DCDB) and recent studies. Notably, within each cancer-specific dataset, only about 5 %–9 % of all combinations were effective, as shown in Table 1. We then

Table 3
Comparative performance metrics of classification models based on LOOCV.

Dataset	Method	Accuracy	Recall	Specificity	Precision	F-Score	MCC	AUC	AUPR	*p*-value
Breast Cancer Dataset	LR	0.852469	0.391304	0.879169	0.157895	0.225000	0.179672	0.510274	0.151638	0.214986
	SVM	0.833432	0.434783	0.856514	0.149254	0.222222	0.180989	0.536467	0.113631	0.208170
	RF	0.729923	0.500000	0.743235	0.101322	0.168498	0.124604	0.633571	0.100556	0.143888
	SGDC	0.856038	0.282609	0.889239	0.128713	0.176871	0.120209	0.562577	0.098605	0.177259
	NLLSS	**0.867936**	0.456522	**0.891756**	0.196262	0.274510	0.237662	0.600473	0.123428	0.318696
	RW	0.842951	0.478261	0.864065	0.169231	0.250000	0.215334	0.670082	0.178663	0.323176
	RWRDC	0.853659	**0.73913**	0.860289	**0.234483**	**0.356021**	**0.360848**	**0.858319**	**0.550367**	
Colorectal Cancer Dataset	LR	0.858025	0.153846	0.919463	0.142857	0.148148	0.070881	0.446373	0.0846	0.045641
	SVM	0.845679	0.192308	0.902685	0.147059	0.166667	0.084206	0.493676	0.106812	0.049304
	RF	0.888889	0.038462	0.963087	0.083333	0.052632	0.002228	0.434757	0.070955	0.043301
	SGDC	**0.895062**	0.038462	**0.969799**	0.100000	0.055556	0.012975	0.49445	0.081007	0.051427
	NLLSS	0.759259	0.384615	0.791946	0.138889	0.204082	0.115378	0.391843	0.085383	0.027443
	RW	0.635802	0.692308	0.630872	0.140625	0.233766	0.1796	0.600929	0.122275	0.042143
	RWRDC	0.802469	**1.000000**	0.785235	**0.288889**	**0.448276**	**0.476283**	**0.933918**	**0.774803**	
Lung Cancer Dataset	LR	0.506944	0.863636	0.477443	0.120253	0.211111	0.182057	0.554511	0.092115	0.000337
	SVM	0.826388	0.636363	0.842105	0.250000	0.358974	0.321120	0.732826	0.414010	0.004538
	RF	0.560763	**0.886363**	0.533834	0.135888	0.235649	0.223226	0.680023	0.116705	0.001107
	SGDC	0.821181	0.136343	0.877820	0.084507	0.104348	0.011460	0.447133	0.069816	0.000969
	NLLSS	0.784722	0.772727	0.785714	0.229729	0.354166	0.339474	0.761876	0.241161	0.005681
	RW	0.776041	0.681818	0.783834	0.206896	0.317460	0.284984	0.714798	0.148580	0.003050
	RWRDC	**0.975694**	0.681818	**1.000000**	**0.909604**	**0.810810**	**0.815067**	**0.909603**	**0.747903**	

Table 4
Case study of breast cancer.

Rank	Drug1	Drug2	Evidence	Rank	Drug1	Drug2	Evidence
1	Arzoxifene	LG100268	DCDB	43	Cyclophosphamide	Trastuzumab	DCDB
2	Decitabine	Depsipeptide	DCDB	44	Bevacizumab	Trastuzumab	DCDB
3	Taxane	Gefitinib	DCDB	45	Trastuzumab	Tamoxifen	DCDB
4	PD98059	Gefitinib	DCDB	46	Paclitaxel	Trastuzumab	DCDB
5	Fulvestrant	Tipifarnib	DCDB	47	PD98059	Taxane	Null
6	Olaparib	Cediranib	DCDB	48	Tipifarnib	Anastrozole	Null
7	Vinorelbine	Neratinib	DCDB	49	Paclitaxel	Cediranib	Null
8	Cisplatin	Gemcitabine	DCDB	50	Neratinib	Trastuzumab	[20]
9	Doxorubicin	Mitomycin	DCDB	51	Fulvestrant	Goserelin acetate	[21]
10	Doxorubicin	Metformin	DCDB	52	Letrozole	Exemestane	Null
11	Methylseleninic acid	Tamoxifen	DCDB	53	Metformin	Mitomycin	Null
12	Cyclophosphamide	Pemetrexed	DCDB	54	Cisplatin	Docetaxel	[22]
13	Sorafenib	Capecitabine	DCDB	55	Everolimus	Tamoxifen	[23]
14	Ixabepilone	Capecitabine	DCDB	56	Anastrozole	Tamoxifen	Null
15	Paclitaxel	Tubacin	DCDB	57	Methylseleninic acid	Exemestane	Null
16	Paclitaxel	Trabectedin	DCDB	58	Methylseleninic acid	Goserelin acetate	Null
17	Paclitaxel	Toremifene	DCDB	59	Everolimus	Lapatinib	Null
18	Epirubicin	Trastuzumab	DCDB	60	Cisplatin	Trastuzumab	[24]
19	Trastuzumab	Bortezomib	DCDB	61	Ixabepilone	Sorafenib	Null
20	Carboplatin	Trastuzumab	DCDB	62	Docetaxel	Mitomycin	Null
21	Fulvestrant	Anastrozole	DCDB	63	Metformin	Docetaxel	[25]
22	Anastrozole	Goserelin acetate	DCDB	64	Pemetrexed	Capecitabine	[26]
23	Everolimus	Exemestane	DCDB	65	Paclitaxel	Pemetrexed	[27]
24	Everolimus	Letrozole	DCDB	66	Trastuzumab	Mitomycin	Null
25	Tamoxifen	Exemestane	DCDB	67	Metformin	Trastuzumab	[28]
26	Tamoxifen	Goserelin acetate	DCDB	68	Trastuzumab	Pemetrexed	Null
27	Lapatinib	Letrozole	DCDB	69	Methylseleninic acid	Trastuzumab	Null
28	Paclitaxel	Olaparib	DCDB	70	Goserelin acetate	Exemestane	[29]
29	Vinorelbine	Trastuzumab	DCDB	71	Trabectedin	Tubacin	Null
30	Docetaxel	Gemcitabine	DCDB	72	Toremifene	Tubacin	Null
31	Bevacizumab	Lapatinib	DCDB	73	Toremifene	Trabectedin	Null
32	Doxorubicin	Docetaxel	DCDB	74	Paclitaxel	Capecitabine	[30]
33	Gemcitabine	Trastuzumab	DCDB	75	Sorafenib	Lapatinib	Null
34	Lapatinib	Capecitabine	DCDB	76	Ixabepilone	Lapatinib	Null
35	Bevacizumab	Capecitabine	DCDB	77	Bevacizumab	Sorafenib	Null
36	Cyclophosphamide	Capecitabine	DCDB	78	Ixabepilone	Bevacizumab	[31]
37	Docetaxel	Capecitabine	DCDB	79	Ixabepilone	Cyclophosphamide	[32]
38	Paclitaxel	Bevacizumab	DCDB	80	Cyclophosphamide	Sorafenib	Null
39	Paclitaxel	Lapatinib	DCDB	81	Sorafenib	Docetaxel	Null
40	Paclitaxel	Cyclophosphamide	DCDB	82	Ixabepilone	Docetaxel	Null
41	Docetaxel	Trastuzumab	DCDB	83	Doxorubicin	Gemcitabine	[33]
42	Doxorubicin	Trastuzumab	DCDB	84	Trastuzumab	Capecitabine	[34]

closely examined the top 10 % of these predictions to focus on the most promising drug combinations. Table 4 outlines these top-tier predictions, with instances labeled as 'Null' indicating a lack of corroborative documentation in the prevailing scientific corpus.

Breast cancer is the most frequently diagnosed cancer and the leading cause of cancer-related mortality in females worldwide [35], comprising 22 % of all cancers in women [36]. We used RWRDC to predict potential effective drug combinations. As a result, 100 % of the drug-drug pairs ranked in the top 5 % (a total of 42 pairs) and 73 % of those in the top 10 % (a total of 84 pairs) of the dataset were validated. When we considered the drug-drug pairs with the top 10 % highest predicted effective scores as the predicted result, all the known effective drug combinations contained in the dataset were accurately validated. Although some predictions have not been validated in databases and literature, there is still a strong possibility that predicted effective drug combinations may prove effective. For example, everolimus-lapatinib was predicted to be 59th by RWRDC and there was no direct evidence that it was an effective drug combination. However, reference [37] indicated that they were conducting a phase II pilot study involving some patients with HER-2 positive metastatic breast cancer (MBC). Additionally, the combination of sorafenib and lapatinib ranked 75th by RWRDC. Reference [38] indicated that both sorafenib and lapatinib alone were effective in the treatment of breast cancer, suggesting that a combination of these two agents may be a promising therapeutic option for the treatment of breast cancer. Furthermore, sorafenib and docetaxel ranked 81st in the prediction results, despite the absence of specific literature indicating that they could be combined to treat breast cancer. A phase II trial of sorafenib combined with docetaxel in the treatment of HER2-negative MBC was conducted [39]. The existing data do not conclusively demonstrate that sorafenib and docetaxel can treat breast cancer; however, with the ongoing development of the trial, clearer conclusions are expected soon.

4. Discussion and conclusion

In recent years, researchers have increasingly recognized the importance of graph methods in bioinformatics research. Graphs not only systematically explore the molecular complexities of specific diseases, identify disease modules and signaling pathways, but also reveal the potential mechanisms of drug action [40]. Graph methods, focusing on interactions between nodes, have been applied in recent studies on drug combinations [41].

The Random Walk with Restart for Drug Combination (RWRDC) model proposed in this study, as a graph-based method, represents a significant advancement in predictive analysis, especially in the treatment of cancer. This model uses the principles of graph theory and employs a random walk strategy to analyze the potential effects of drug combinations, providing new perspectives and tools for the screening and optimization of drug combinations. This model extends the classical random walk within a stochastic process framework, and we have proven the convergence of the algorithm, providing a solid mathematical foundation. This progress means that we can find exact solutions to specific problems instead of relying on iterative numerical solutions, greatly reducing computation time and increasing accuracy for large-scale problems. The core of the model is to achieve a steady state in the random process, which helps calculate the effective scores of drug combinations, enabling the RWRDC model to identify promising pairs for targeted disease treatment.

Extensive empirical validation and cross-validation methods have demonstrated the RWRDC model's strong predictive power in forecasting drug combinations for cancer and other complex diseases. Additionally, numerous reports have confirmed the efficacy of natural drug combinations, emphasizing the critical role of natural compounds in drug synergy [42–44]. Herbal medicines typically contain various natural active ingredients, and their therapeutic effects are achieved through a "multi-component, multi-target" mechanism. However, as the characteristics of many herbal medicines have not been systematically collected and organized, the RWRDC semi-supervised learning framework, which does not rely on additional features, is particularly suitable for scenarios with limited data availability.

Despite the significant effectiveness of the RWRDC model, there are limitations. Firstly, as a semi-supervised learning framework, it has not yet integrated additional data features, such as chemical structure similarities between drugs, drug-pathway interactions, and drug-target interactions, which could significantly enhance the model's accuracy [45, 46]. Secondly, the current research focuses on effective combinations of two drugs, but the model has the potential to expand to multiple drug combinations. Future iterations of the model could adopt a more complex random walk framework to predict such combinations. Thirdly, the lack of a negative dataset is a challenge in computational modeling, as successful drug combinations are more frequently reported while negative combinations receive less attention. Lastly, the issue of drug combinations being effective or ineffective is often related to drug dosage, which warrants further exploration [47]. Although some studies have considered the impact of drug dosage, this issue requires more in-depth research to achieve the goal of precision medicine.

In summary, the RWRDC model is a key resource for identifying potential effective drug combinations and provides profound insights into the underlying mechanisms of drug interactions. It not only paves the way for the development of new treatment strategies but also broadens the horizons of medical research in effective drug combination prediction.

Ethics approval and consent to participate

Not applicable.

Consent to publish

Not applicable.

Availability of data and materials

The code of RWRDC is freely available at https://github.com/wangqi27/ERWR.

Competing interests

The authors declare that they have no competing interests.

CRediT authorship contribution statement

Qi Wang: Writing – review & editing, Writing – original draft, Visualization, Validation, Software, Methodology, Investigation, Formal analysis, Conceptualization. **Xiya Liu:** Software, Validation, Writing – review & editing. **Guiying Yan:** Writing – review & editing, Supervision, Funding acquisition, Conceptualization.

Declaration of competing interest

The 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.

Acknowledgements

This work was supported by the National Natural Science Foundation of China, China (No.12231018). And the authors would like to acknowledge Dr. Yaqi Zhang for technical support.

List of abbreviations

AUC	Area Under the Receiver Operating Characteristic Curve
ROC	Receiver Operating Characteristic
AUPR	Area Under the Precision-Recall Curve
MCC	Matthews Correlation Coefficient
TP	True Positive
TN	True Negative
FN	False Negative
FP	False Positive
LOOCV	Leave-One-Out Cross-Validation
DCDB	Drug Combination Database
MBC	Metastatic Breast Cancer
WHO	World Health Organization
PS	Positive Samples (effective combinations)
CS	Candidate Samples (unknown combinations)
NS	Negative Samples (non-effective combinations in candidate samples)
RWRDC	Random Walk with Restart for Drug Combination
NLLSS	Network-based Laplacian regularized Least Square Synergistic drug combination prediction
RW	Random Walk
LR	Logistic Regression
SVM	Support Vector Machine
RF	Random Forest
SGDC	Stochastic Gradient Descent Classifier

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.synbio.2024.09.003.

References

[1] Liu J, Gefen O, Ronin I, et al. Effect of tolerance on the evolution of antibiotic resistance under drug combinations. Science 2020;367:200–4.
[2] Liu Y, Wei Q, Yu G, Gai W, Li Y, Chen X. DCDB 2.0: a major update of the drug combination database. Database 2014;2014:bau124.
[3] Jaaks P, Coker EA, Vis DJ, et al. Effective drug combinations in breast, colon, and pancreatic cancer cells. Nature 2022;603:166–73.
[4] Gibas KM, Kelly SG, Arribas JR, Cahn P, Orkin C, Daar ES, Sax PE, Taiwo BO. Two-drug regimens for HIV treatment. Lancet HIV 2022;9(12):e868–83.
[5] Tamma PD, Cosgrove SE, Maragakis LL. Combination therapy for treatment of infections with gram-negative bacteria. Clin Microbiol Rev 2012;25(3):450–70.
[6] Zhao Y, Yin J, Zhang L, Zhang Y, Chen X. Drug-drug interaction prediction: databases, web servers, and computational models. Briefings Bioinf 2024;25(1):bbad445.
[7] Pan Y, Ren H, Lan L, Li Y, Huang T. Review of predicting synergistic drug combinations. Life 2023;13(9):1878.
[8] Tonekaboni SAM, Ghoraie LS, Manem VSK, Haibe-Kains B. Predictive approaches for drug combination discovery in cancer. Briefings Bioinf 2018;19(2):263–76.
[9] Chen X, Ren B, Chen M, Wang Q, Zhang L, Yan G. NLLSS: predicting synergistic drug combinations based on semi-supervised learning. PLoS Comput Biol 2016;12(7):e1004975.
[10] Hansen PW, Clemmensen L, Sehested TS, et al. Identifying drug-drug interactions by data mining: a pilot study of warfarin-associated drug interactions. Circulation: Cardiovascular Quality and Outcomes 2016;9(6):621–8.
[11] Cüvitoğlu A, Işik Z. Classification of effects of drug combinations with support vector machines. In: 2017 25th signal processing and communications applications conference (SIU). IEEE; 2017. p. 1–4.
[12] Huang H, Zhang P, Qu X, et al. Systematic prediction of drug combinations based on clinical side-effects. Sci Rep 2014. *Article*.
[13] Xu Q, Xiong Y, Dai H, et al. Pdc-sgb: prediction of effective drug combinations using a stochastic gradient boosting algorithm. J Theor Biol 2017;417:1–7.
[14] Han C-D, Wang C-C, Huang L, Chen X. MCFF-MTDDI: multi-channel feature fusion for multi-typed drug–drug interaction prediction. Briefings Bioinf 2023;24(4):bbad215.
[15] Barabási AL, Gulbahce N, Loscalzo J. Network medicine: a network-based approach to human disease. Nat Rev Genet 2011;12(1):56–68.
[16] Hopkins AL. Network pharmacology: the next paradigm in drug discovery. Nat Chem Biol 2008;4(11):682–90.
[17] Jemal A, Center MM, DeSantis C, Ward EM. Global patterns of cancer incidence and mortality rates and trends. Cancer Epidemiol Biomarkers Prev 2010;19(8):1893–907.
[18] Quan H, Sundararajan V, Halfon P, Fong A, Burnand B, Luthi J-C, et al. Coding algorithms for defining comorbidities in ICD-9-CM and ICD-10 administrative data. Med Care 2005;43(11):1130–9.
[19] Tong H, Faloutsos C, Pan J. Random walk with restart: fast solutions and applications. Knowl Inf Syst 2008;14:327–46.
[20] Chen X, Liu M-X, Yan G-Y. Drug-target interaction prediction by random walk on the heterogeneous network. Mol Biosyst 2012;8(7):1970–8.
[21] Li Y, Patra JC. Genome-wide inferring gene-phenotype relationship by walking on the heterogeneous network. Bioinformatics 2010;26(9):1219–24.
[22] Wang L. Docetaxel-cisplatin combined chemotherapy in advanced breast cancer. Chin J Cancer Res 2007;14:1102–3.
[23] Bachelot T, Bourgier C, Cropet C, Ray-Coquard I, Ferrero J-M, Freyer G, Co-authors. Randomized phase II trial of everolimus in combination with tamoxifen in patients with hormone receptor-positive, human epidermal growth factor receptor 2-negative metastatic breast cancer with prior exposure to aromatase inhibitors: a GINECO study. J Clin Oncol 2012;30(22):2718–24.
[24] Pegram MD, Slamon DJ. Combination therapy with trastuzumab (Herceptin) and cisplatin for chemoresistant metastatic breast cancer: evidence for receptor-enhanced chemosensitivity. Semin Oncol 1999;26(4 Suppl 12):89–95.
[25] Yi Z, Qingli Z, Ji M. Proliferation and apoptosis effect of combining metformin and docetaxel on breast cancer cells. Journal of Modern Oncology 2018.
[26] Xiaorong Y, Lingru M, Huanxin Y, et al. Clinical observation of pemetrexed combined with capecitabine in the treatment of recurrent and metastatic breast cancer. Journal of Basic and Clinical Oncology 2015.
[27] Tas F, Duranyildiz D, Oguz H, Camlica H, Yasasever V, Topuz E. Circulating levels of vascular endothelial growth factor (VEGF), matrix metalloproteinase-3 (MMP-3), and BCL-2 in malignant melanoma. Med Oncol 2008;25(4):431–6.
[28] Thor A, Fan Z, Yang X, Esteva F, Jones F, Edgerton S, & Co-authors. Metformin enhances trastuzumab efficacy and overcomes resistance in HER2 type breast cancer cells, vol. 2009. Poster Session Abstracts; 2009. p. 1133.
[29] Yang L, Xin T, Rong WU. Efficacy of combined therapy of goserelin and exemestane on the advanced and metastatic breast cancer patients in the premenopausal and perimenopausal. Practical Pharmacy and Clinical Remedies 2014.
[30] Gradishar WJ, Meza LA, Amin B, Samid D, Hill T, Chen Y-M, Co-authors. Capecitabine plus paclitaxel as front-line combination therapy for metastatic breast cancer: a multicenter phase II study. J Clin Oncol 2004;22(12):2321–7.
[31] Verrill M. A trial looking at either ixabepilone or paclitaxel with bevacizumab for breast cancer that has come back (CA163115). Retrieved from, https://www.cancerresearchuk.org/about-cancer/find-a-clinical-trial/a-trial-looking-at-either-ixabepilone-or-paclitaxel-chemotherapy-in-combination-with-bevacizumab-for-breast-cancer-that-has-come-back-locally-or-has-spread#undefined; 2013.
[32] Yardley DA, Peacock NW, Shastry M, Burris HA, Bechhold RG, Hendricks CB, Co-authors. A phase II trial of ixabepilone and cyclophosphamide as neoadjuvant therapy for patients with HER2-negative breast cancer: Correlation of pathologic complete response with the 21-gene recurrence score. Breast Cancer Res Treat 2015;154(2):299–308.
[33] Pan G-Q, Ren H-Z, Zhang S-F, Wang X-M, Wen J-F. Expression of semaphorin 5A and its receptor plexin B3 contributes to invasion and metastasis of gastric carcinoma. World J Gastroenterol 2009;15(22):2800–4.
[34] Bartsch R, Wenzel C, Altorjai G, Pluschnig U, Rudas M, Mader RM, Co-authors. Capecitabine and trastuzumab in heavily pretreated metastatic breast cancer. J Clin Oncol 2007;25(25):3853–8.
[35] Zheng R-R, Hu W, Sui C-G, Ma N, Jiang Y-H. Effects of doxorubicin and gemcitabine on the induction of apoptosis in breast cancer cells. Oncol Rep 2014;32(6):2719–25.
[36] Karagoz K, Sinha R, Arga KY. Triple negative breast cancer: a multi-omics network discovery strategy for candidate targets and driving pathways. OMICS 2015;19(2):115–30.
[37] Barr JA, Sharma P, Fabian CJ, Yeh H, Baccaray S, Springer M, Co-authors. Phase II trial of lapatinib and everolimus for HER2 positive metastatic breast cancer. Ongoing Clinical Trials. Abstracts: Thirty-Eighth Annual CTRC-AACR San Antonio Breast Cancer Symposium; December 8-12, 2015; San Antonio, TX: American Association for Cancer Research; 02152016 2016. OT3-01-12-OT3-01-12.
[38] Kacan T, Altun A, Altun GG, Kacan SB, Sarac B, Seker MM, Co-authors. Investigation of antitumor effects of sorafenib and lapatinib alone and in combination on MCF-7 breast cancer cells. Asian Pac J Cancer Prev APJCP 2014;15(7):3185–9.
[39] Marme F, Gerber B, Schmidt M, Moebus VJ, Foerster FG, Grischke E-M, Co-authors. Sorafenib (SOR) plus docetaxel (DOC) as first-line therapy in patients with HER2-negative metastatic breast cancer (MBC): a randomized, placebo-controlled phase II trial. J Clin Oncol 2014;32(15_suppl):1072.

[40] Cheng F, Kovács IA, Barabási AL. Network-based prediction of drug combinations. Nat Commun 2019;10(1):1–11.
[41] Paltun BG, Kaski S, Mamitsuka H. Machine learning approaches for drug combination therapies. Briefings Bioinf 2021;22(6):bbab293.
[42] Ma XH, Zheng CJ, Han LY, et al. Synergistic therapeutic actions of herbal ingredients and their mechanisms from molecular interaction and network perspectives. Drug Discov Today 2009;14:579–88.
[43] Yuan R, Lin Y. Traditional Chinese medicine: an approach to scientific proof and clinical validation. Pharmacol Therapeut 2000;86(2):191–8.
[44] Koeberle A, Werz O. Multi-target approach for natural products in inflammation. Drug Discov Today 2014;19(11):1871–82.
[45] Adam G, Rampášek L, Saffkhani Z, et al. Machine learning approaches to drug response prediction: challenges and recent progress. npj Precis Oncol 2020;4(1):1–10.
[46] Wang C-C, Zhao Y, Chen X. Drug-pathway association prediction: from experimental results to computational models. Briefings Bioinf 2021;22(3):bbaa061.
[47] Julkunen H, Cichonska A, Gautam P, Szedmak S, Douat J, Pahikkala T, Aittokallio T, Rousu J. Leveraging multi-way interactions for systematic prediction of pre-clinical drug combination effects. Nat Commun 2020;11:6136.