scientific reports

Check for updates

OPEN

NMGMDA: a computational model for predicting potential microbe–drug associations based on minimize matrix nuclear norm and graph attention network

Mingmin Liang[1], Xianzhi Liu[1], Qijia Chen[1✉], Bin Zeng[1✉] & Lei Wang[1,2✉]

The prediction of potential microbe–drug associations is of great value for drug research and development, especially, methods, based on deep learning, have been achieved significant improvement in bio-medicine. In this manuscript, we proposed a novel computational model named NMGMDA based on the nuclear norm minimization and graph attention network to infer latent microbe–drug associations. Firstly, we created a heterogeneous microbe–drug network in NMGMDA by fusing the drug and microbe similarities with the established drug–microbe associations. After this, by using GAT and NNM to calculate the predict scores. Lastly, we created a fivefold cross validation framework to assess the new model NMGMDA's progressiveness. According to the simulation results, NMGMDA outperforms some of the most advanced methods, with a reliable AUC of 0.9946 on both MDAD and aBioflm databases. Furthermore, case studies on Ciprofloxacin, Moxifoxacin, HIV-1 and Mycobacterium tuberculosis were carried out in order to assess the effectiveness of NMGMDA even more. The experimental results demonstrated that, following the removal of known correlations from the database, 16 and 14 medications as well as 19 and 17 microbes in the top 20 predictions were validated by pertinent literature. This demonstrates the potential of our new model, NMGMDA, to reach acceptable prediction performance.

Microorganisms are a class of widely dispersed germs that include bacteria, viruses, fungus, and other species that are both helpful and hazardous to humans[1,2]. Numerous human organs contain and are covered in human microbes[3]. In addition to promoting food absorption and maintaining intestinal health by managing the balance of the gut microbiota, they can control the host's mucosal and systemic immune systems[4,5]. In the intestinal environment, these bacteria depend on one another and benefit one another. When the microbiota is out of balance, several diseases, include obesity[6], inflammatory bowel disease[7], and cancer[8] can result. Additionally, numerous studies have demonstrated that while utilizing pharmaceuticals to cure diseases, there is a definite influence between bacteria and drugs[1,9,10]. Therefore, understanding the relationship between microbes and medications becomes essential for the treatment of disease.

Humans have discovered certain relationships between drugs and microbes through investigations into biology, but because biological experiments demand a substantial amount of human, material, and time resources, their further advancement may be constrained. To address the limitations of biological studies, an increasing number of computational methods have been presented during the past few years due to the rapid development of relevant research tools. These methods aim to anticipate the relationship between drugs and microbes[11]. In parallel, databases of microbe–drug associations that have undergone experimental validation, such as MDAD[12] and aBioflm[13], have also been established. For instance, Zhu et al.[14] presented HMDAKATZ, which uses the KATZ measure to identify microbe–drug associations. By integrating a network embedding approach with microbe–drug association prediction, Long et al.[15] introduced the HNERMDA method. To predict probable microbe–drug associations, Ma et al.[16] introduced the generalized Matrix decomposition method WHGMF based on weighted hypergraph learning. To infer new microbe–drug relationships, Yang et al.[17] suggested the

---

[1]School of Information Engineering, Hunan Vocational College of Electronic and Technology, Changsha 410000, China. [2]Big Data Innovation and Entrepreneurship Education Center of Hunan Province, Changsha University, Changsha 410022, China. ✉email: 752319383@qq.com; 13974880055@139.com; wanglei@xtu.edu.cn

nature portfolio

1

multi-core fusion model MKGNN based on Graph Convolutional Network(GCN). A deep neural network-based prediction model for microbe–drug associations called NNAN was created by Zhu et al.[18] A contrastive learning model called SCSMDA was created by Tian et al.[19] to forecast the connection between microbes and drugs. In order to anticipate probable microbe–drug correlations, Tan et al.[20] developed a computation technique termed GSAMDA based on the graph attention network and the sparse autoencoder. Yang et al.[21] suggest a model, called MKGCN, for inferring microbe–drug associations based on Multiple Kernel Fusion on Graph Convolutional Network. Ma et al.[22] designed a microbe–drug prediction model based on graph attention network (GAT) and convolutional neural networks (CNN).

As mentioned above, it is easy to know that these neural network-based methods are frequently used in hiding random association prediction works, and among them, CNN-based approaches adopt the method of parameter sharing to effectively prevent overfitting, however, the pooling layer will lose a significant amount of important data during processing. As for the GCN-based approaches, although the non-matrix organized data will be more applicable, however, the scalability and flexibility are still quite limited. As for the GAT-based methods, although the clustering performance of graph neural networks can be significantly improved, but the clustering of higher-order neighborhoods is still a challenging task. Hence, it is clear that better prediction results can be obtained by combining these above prediction methods organically.

In this study, we introduced a novel calculating approach called NMGMDA to predict latent associations between microbes and drugs, which is based on the nuclear norm minimization[23] and the graph attention network[24]. Figure 1 depicts the NMGMDA structure. These are our primary contributions, in brief:

- A novel heterogeneous network made up of microbes and drugs has been created by combining the microbe similarity network, drug similarity network, and existing microbe–drug relationships.
- To get projected scores for potential microbe–drug associations, we used both the nuclear norm minimization (NNM) approach and the GAT-based auto-encoder. And then weighted averaged these two predicted scores to get the final results.
- Experimental results and case studies demonstrated the significant prediction performance of NMGMDA on both the MDAD and the aBioflm Databases.

Materials and methods
Data sources
In this study, we assessed NMGMDA on the following two databases in order to show its efficacy.

MDAD database is a database of microbe–drug associations that was assembled and arranged by Sun et al.[12] in 2018 from a variety of drug-related databases, including TTD and DrugBank, as well as a substantial body of literature. After superfluous data is eliminated, 1373 drugs and 173 microbes were found to have 2470 microbe–drug associations.
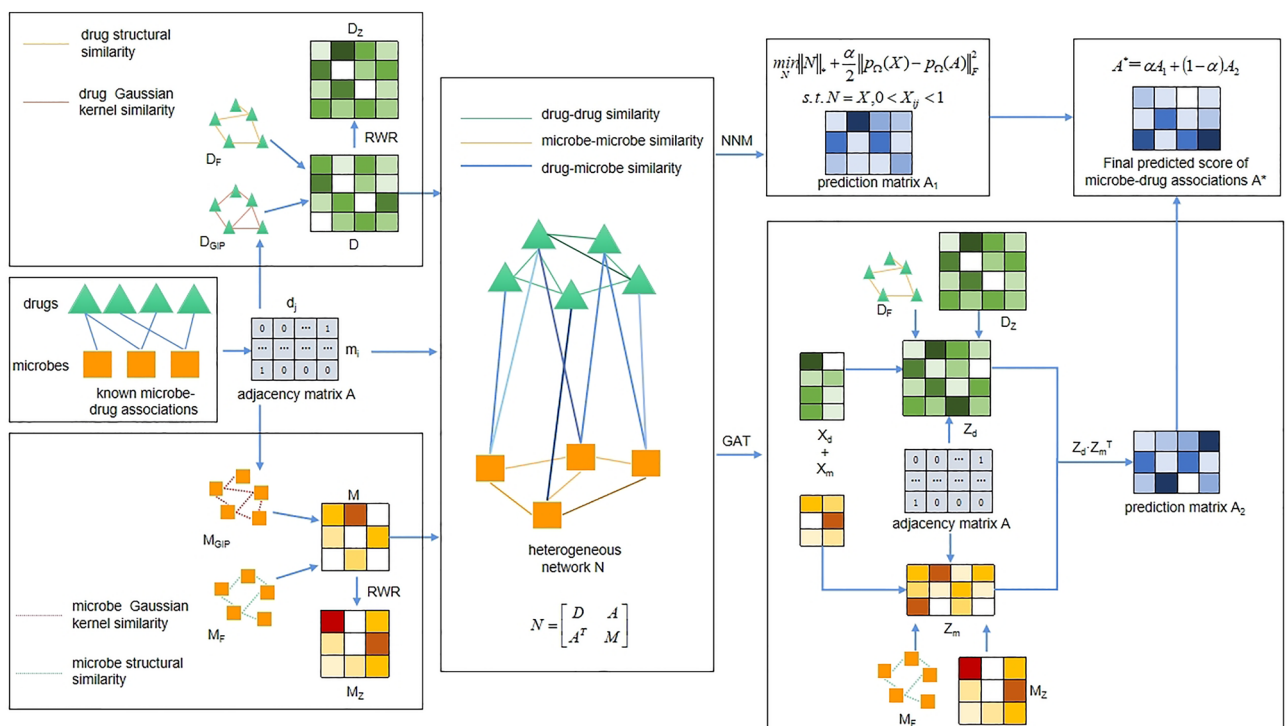

Figure 1. The overall architecture of NMGMDA.

ABiofilm database was created by Rajput et al.[13], which includes 5027 antifungal drugs that target 140 microbes that were identified between 1988 and 2017. Following the removal of redundant data, 140 microbes and 1720 drugs were included in 2884 microbe–drug associations.

Table 1 provides specific statistics of microbes-drugs associations in the MDAD and aBioflm.

Methods

Microbe–drug adjacency matrix

We initially create an adjacency matrix $A \in R^{n_d \times n_m}$, where n_d and n_m represent the number of drugs and microbes, respectively, based on these microbe–drug associations. A_{ij} equals 1 if there is a known relationship between the drug d_i and microbe m_j, else it equals 0.

$$A_{ij} = \begin{cases} 1, & \text{if } d_i \text{ associats with } m_j \\ 0, & \text{otherwise} \end{cases} \tag{1}$$

Drug/microbe Gaussian kernel similarity

The following formula will be used to determine the Gaussian kernel similarity $D_{GIP}(d_i, d_j) \in R^{n_d \times n_d}$ between d_i and d_j, assuming that d_i and d_j are two drugs.

$$D_{GIP} = exp\left(-\gamma_d \|A(d_i) - A(d_j)\|^2\right) \tag{2}$$

where $\|A(d_i) - A(d_j)\|$ is the Euclidean distance between two drugs. Since γ_d is a regular parameter, it is easier to group together similar feature points the greater γ_d. And the definition of γ_d is as follows:

$$\gamma_d = 1 / \left(\frac{1}{n_d} \sum_{i=1}^{n_d} \|A(d_i)\|^2\right) \tag{3}$$

Similarly, we would calculate the Gaussian kernel similarity $M_{GIP}(m_i, m_j) \in R^{n_m \times n_m}$ between two microbes:

$$M_{GIP} = exp\left(-\gamma_m \|A(m_i) - A(m_j)\|^2\right) \tag{4}$$

$$\gamma_m = 1 / \left(\frac{1}{n_m} \sum_{i=1}^{n_m} \|A(m_i)\|^2\right) \tag{5}$$

Microbe/Drug functional similarity

In the STRING[25] database, we can find many gene functional networks connected to microbes. A matrix $M_F \in R^{n_m \times n_m}$ can be produced by the Kamneva[26] tool, which determines microbe functional similarity based on microbial gene families.

The SIMCOMP2 tool[27] uses the chemical and molecular formula structures of drugs to determine how similar their structures are. To create a drug functional similarity matrix $D_F \in R^{n_d \times n_d}$, we adopt the similarity scores.

Drug/microbe integrated similarities

It is important to note that not every drug can determine functional similarity. As a result, using the drug structural similarity and the drug Gaussian kernel similarity, we were able to construct a new matrix $D \in R^{n_d \times n_d}$ of integrated drug similarities.

$$D = \begin{cases} (D_{GIP} + D_F)/2, & if D_F \neq 0 \\ D_{GIP}, & if D_F = 0 \end{cases} \tag{6}$$

where D_{GIP} is the drug Gaussian kernel similarity, and D_F is the drug functional similarity.

Similarly, the microbe integrated similarities matrix $M \in R^{n_m \times n_m}$ was calculated as follows:

$$M = \begin{cases} (M_{GIP} + M_F)/2, & if M_F \neq 0 \\ M_{GIP}, & if M_F = 0 \end{cases} \tag{7}$$

where M_{GIP} is the drug Gaussian kernel similarity, and M_F is the drug functional similarity.

Database	Microbes	Drugs	Associations
MDAD	173	1373	2470
aBioflm	140	1720	2884

Table 1. The specific statistics of microbes-drugs associations in the MDAD and aBioflm.

Constructing the heterogeneous network *N*

The microbe–drug adjacency matrix, drug integrated similarities matrix and microbe integrated similarities matrix can be joined together to form a whole matrix $N \in R^{(n_d+n_m)\times(n_d+n_m)}$:

$$N = \begin{bmatrix} D & A \\ A^T & M \end{bmatrix} \tag{8}$$

where A^T represents $A\prime$s transposition.

Predicting microbe–drug associations by NNM

Currently, the convex optimization model includes nuclear norms, which are applied in many fields[28]. It has a globally optimal solution[11]. Therefore, the nuclear norm minimization of the heterogeneous network N can be expressed as:

$$\min_N \|N\|_* \quad s.t. N_{ij} = A_{ij}, (i,j) \in \Omega \tag{9}$$

where $\|N\|_*$ represents the nuclear norm of N, Ω is a set of known positions of elements.

We need to add restrictions to the model to make sure that the unknown elements fall within the range [0,1] since predicted scores for microbe–drug associations should be between [0,1]. This forecasting method is:

$$\min_N \|N\|_* s.t. \|p_\Omega(N) - p_\Omega(A)\|_F \le \varepsilon \tag{10}$$

$$\left(p_\Omega(N)\right)_{ij} = \begin{cases} N_{ij}, if (i,j) \in \Omega \\ 0, otherwise \end{cases} \tag{11}$$

They are ε, which stands for measurement noise, $\|\cdot\|_F$, which stands for the Frobenius norm, and p_Ω, which stands for the orthogonal mapping acting on Ω. Then substituting regularized models for inequality constrained models:

$$\min_N \|N\|_* + \frac{\alpha}{2} \|P_\Omega(N) - p_\Omega(A)\|_F^2 \quad s.t. 0 < N_{ij} < 1 \tag{12}$$

where α is a variable that is learnable. The model can be optimized in the manner shown below by introducing the auxiliary matrix X, which was inspired by literature[29]:

$$\min_N \|N\|_* + \frac{\alpha}{2} \|P_\Omega(X) - p_\Omega(A)\|_F^2 \quad s.t. N = X, 0 < X_{ij} < 1 \tag{13}$$

Then, minimize the enhanced Lagrange function to solve the problem:

$$\zeta(X,N,Y,\alpha,\beta) = \|N\|_* + \frac{\alpha}{2} \|P_\Omega(X) - p_\Omega(A)\|_F^2 + T_\gamma\left(Y^T(N-X)\right) + \frac{\beta}{2} \|N-X\|_F^2 \tag{14}$$

where Y is the Lagrange multiplier and $\beta > 0$ is the penalty factor.

Following that, implement iterative solution. The matrix X_{k+1} must first be calculated:

$$X_{k+1} = arg \min_{0 \le X \le 1} \zeta(X, N_k, Y_k, \alpha, \beta)$$

$$= arg \min_{0 \le X \le 1} \frac{\alpha}{2} \|P_\Omega(X) - p_\Omega(A)\|_F^2 + T_\gamma\left(Y^T(N-X)\right) + \frac{\beta}{2} \|N-X\|_F^2 \tag{15}$$

The best answer to the Eq. (15) for $arg \min_{0 \le X \le 1} \zeta(X, N_k, Y_k, \alpha, \beta)$ is X^*. Think of restrictions for the interval [0,1] that:

$$\left(X_{k+1}\right)_{ij} = \begin{cases} 1, X_{ij}^* > 1 \\ X_{ij}^*, 0 \le X_{ij}^* \le 1 \\ 0, X_{ij}^* < 0 \end{cases} \tag{16}$$

Update the matrix N_{k+1} and correct other variables:

$$N_{k+1} = arg \min_N \zeta\left(X_{k+1}, N, Y_k, \alpha, \beta\right)$$

$$= arg \min_N \|N\|_* + \frac{\beta}{2} \left\|N - \left(X_{k+1} - \frac{Y_k}{\beta}\right)\right\|^2 = \vartheta_{\frac{1}{\beta}}\left(X_{k+1} - \frac{Y_k}{\beta}\right) \tag{17}$$

$$\vartheta_\tau(x) = \sum_{i=1}^{\theta_i \ge \tau} (\theta_i - \tau)\mu_i v_i^T \tag{18}$$

where $\vartheta_\tau(x)$ is singular value contraction operator, θ_i is the singular values of X which is larger than τ, while μ_i and v_i are the left and right singular vectors corresponding to θ_i.

We can update the Lagrange multiplier Y_{k+1} as follows by adjusting other variables:

$$Y_{k+1} = Y_k + \beta \left(N_{k+1} - X_{k+1} \right) \tag{19}$$

Finally, the following information can be found in the prediction matrix A_1 for microbe–drug associations:

$$X_k \rightarrow \begin{bmatrix} D_1 & A_1 \\ A_1^T & M_1 \end{bmatrix} \rightarrow A_1 \tag{20}$$

Predicting latent microbe–drug associations by GAT

With the introduction of an attention-based design, the graph spatial network GAT performs node categorization for graph-structured data[24]. To determine the matrix N's structure, we created a GAT model. First determines the attention score between any two nodes in the matrix N:

$$\alpha_{ij} = \frac{exp\left(e_{ij}\right)}{\sum_{k \in N_i} exp(e_{ik})} \tag{21}$$

$$e_{ij} = LeakyRelu\left(a\left[Wh_i || Wh_j\right]\right) \tag{22}$$

$$LeakyRelu(x) = \begin{cases} x, & x > 0 \\ \mu x, & otherwise \end{cases} \tag{23}$$

where N_i stands for the total number of nodes, a is an attention coefficient, W is a learnable linear transformation, and h_i represents the feature vector of the node i, μ is the hypermeter and $||$ denotes the concatenation.

Consequently, each node's ultimate output feature is:

$$h_i = relu\left(\sum_{j \in N_i} \alpha_{ij} Wh_j\right) \tag{24}$$

The activation function, $relu$, is defined as follows:

$$relu(x) = \begin{cases} x, & x > 0 \\ 0, & otherwise \end{cases} \tag{25}$$

A low dimensional structural matrix $X = \begin{bmatrix} X_d \\ X_m \end{bmatrix} \in R^{(n_d+n_m) \times l}$ is produced by substituting N into the previously mentioned GAT model, where X_d and X_m, respectively, stand in for the drug nodes and microbial nodes in N. After a number of testing, we ultimately decided on MSE loss as the loss function for optimizing our model.

An improved random walk with restart (RWR) is implemented on D in response to literature[20], allowing us to obtain a new matrix. Below is how the RWR was described:

$$s_i^{l+1} = \lambda X s_i^l + (1 - \lambda)\varepsilon_i \tag{26}$$

$$\varepsilon_{ij} = \begin{cases} 1, & if\ i = j \\ 0, & otherwise \end{cases} \tag{27}$$

where ε_i is the initial probability vector, X is the matrix of transition probabilities, and λ is the restart probability. Similar to that, we might produce a novel matrix M_Z by using the enhanced RWR on M.

As a result, by combining the drug matrix X_d, D_F, D_Z and adjacency matrix A, influenced by literature[22], we could create a new drug feature matrix Z_d that looked like this:

$$Z_d = [X_d, D_F, A, D_Z, A] \tag{28}$$

Similarly, we could create the following new microbe feature matrix:

$$Z_m = \left[X_m, A^T, M_F, A^T, M_Z\right] \tag{29}$$

Finally, we employ dot product to derive a microbe–drug association predictive score A_2:

$$A_2 = swich\left(Z_d(d_i) \cdot Z_m\left(m_j\right)^T\right) \tag{30}$$

$$swich = xSigmoid(\beta x) \tag{31}$$

where *swich* is an activation function, β, a learnable parameter, which is typically set it to 1, $Z_d(d_i)$ indicates the i_{th} row of Z_d and $Z_m(m_j)$ represents the j_{th} row of Z_m.

Final predicted score of microbe–drug associations

The weighted arithmetic mean approach can be used to combine the prediction matrix A_1 acquired through NNM and the prediction matrix A_2 generated through GAT, resulting in the following final forecast matrix A^* of microbe–drug associations:

$$A^* = \lambda A_1 + (1 - \lambda)A_2 \tag{32}$$

where λ is the weight value.

Experiments and results

In this section, we first carried out sensitive parameter analysis to get the optimum performance out of the model. Then, six state-of-the-art methods would be picked to contrast with NMGMDA. Finally, in order to confirm the validity of our model, we have chosen two typical microbes and drugs, respectively.

Parameter sensitivity analysis

Three pieces make up the NMGMDA model. α and β in formula (14) are two crucial parameters in NNM. Dimension l and learning rate l_r are the two most important factors in GAT. The weight value λ is an important parameter in the final prediction formula (32). In this section, to find the appropriate settings and ensure the independence of the training sets and test sets, we initially Randomly picked 20% of the associations are known and 20% are unknown for the training sets, with the remaining sets being test sets. Next, we utilized fivefold CV experiments with the MDAD database and ensure each of the experiments is independent.

In NNM, we decided to conduct joint tests and altered α and β from {0.1, 1, 10, 100, 1000} and conduct joint experiments. Then, using a fivefold CV experiment, we determined the area under curve (AUC) and the area under the precision-recall curve (AUPR) of these parameter combinations. The findings are displayed in Table 2. Table 2 shows that the AUC and AUPR outcomes obtained by NMGMDA are both at their best when α and β have values of 100 and 1, respectively.

In GAT, we decided to adjust the dimension l changed from {16, 32, 64, 128} and the learning rate l_r changed from {0.1, 0.05, 0.01, 0.005, 0.001}.

Figure 2 makes it clear that no substantial changes to the outcome were caused by changing any particular factors. We choose 32 as the dimension of node topological representation l since it has a little better AUPR value than 64 or 128 dimensions. In line with typical learning models, the learning rate l_r was set at 0.01.

Finally, the results are displayed in Fig. 3 for parameter λ in formula (32), where we estimate the impact of the λ altered from {0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9} for the fivefold on MDAD. Which makes it clear that NMGMDA, with λ set to 0.7, may get the maximum AUC and AUPR values.

After comparing the performance on different hyperparameters by testing, the final parameters we selected are $\alpha = 100$, $\beta = 1$, $l = 32$, $l_r = 0.01$ and $\lambda = 0.7$.

Comparison with advanced methods

In this case, taking into account the dearth of microbial drug association prediction methods, we would first contrast NMGMDA with a few standard approaches for link prediction issues, such as HMDAKATZ[14], HMDA-Pred[30], LAGCN[31], MNNMAD[32] and GSAMDA[20], etc.

Here, considering the limited availability of microbial drug association prediction methods, we would first compare NMGMDA with some representative methods for link prediction problems such as HMDAKATZ[14], HMDA-Pred[30], LAGCN[31], MNNMAD[32] and GSAMDA[20], etc. One of them, HMDAKATZ, predicted the association between microbes and drugs using the KATZ algorithm as a foundation. For the prediction of microbe–disease associations, HMDA-Pred is a novel computer model based on multi-data integration and network consistency projection. LAGCN is a complete end-to-end graph based deep learning method, which forecast the associations between drugs and diseases. By using a Matrix Nuclear Norm approach on data on known microbes and diseases, MNNMAD is a method for predicting microbe–disease relationships. Based on

		Average AUC on MDAD				Average AUPR on MDAD			
β	0.1	0.98245	0.99329	0.99331	0.99331	0.67785	0.88056	0.88726	0.88701
	1	0.99331	0.99443	0.99468	0.99467	0.88849	0.89842	0.90734	0.90719
	10	0.99259	0.99236	0.99367	0.99302	0.86172	0.87312	0.88416	0.88725
	100	0.99308	0.99309	0.99309	0.99309	0.87951	0.87911	0.86045	0.85653
		0.1	1	10	100	0.1	1	10	100
		α							

Table 2. The AUC and AUPR values on different α and β on MDAD database.

(a) Effect of the dimension of node topological representation
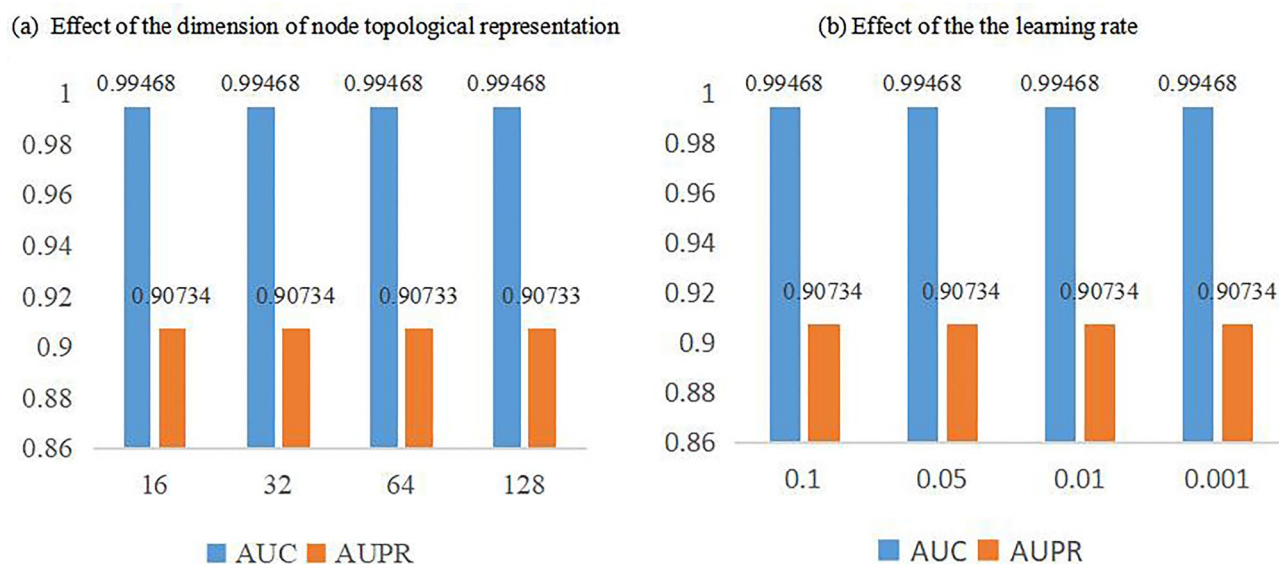

(b) Effect of the the learning rate

Figure 2. The AUC and AUPR values on different dimension of node topological representation and learning rate on MDAD database.
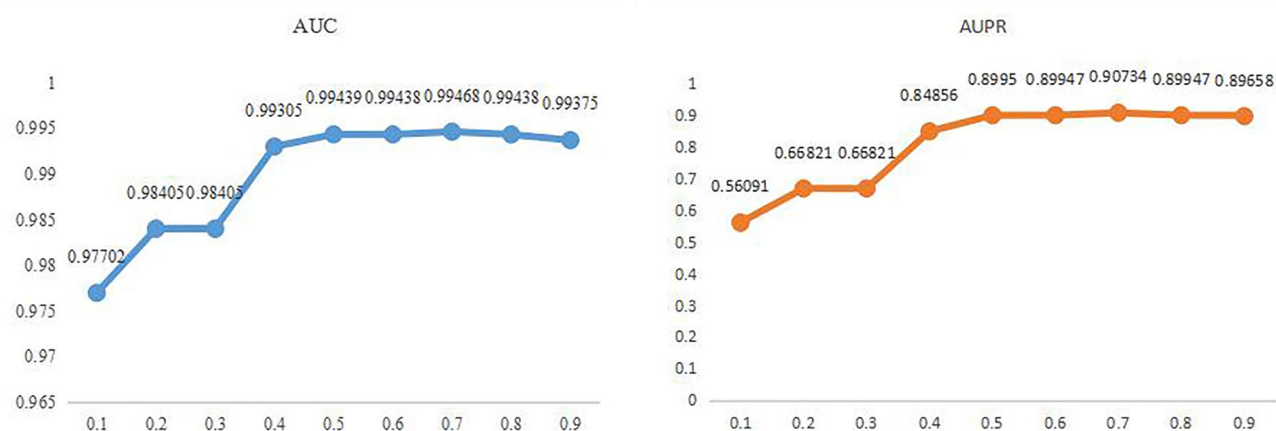
AUC

AUPR

Figure 3. The AUC and AUPR values on different weight value on MDAD database.

a graph attention network and sparse auto-encoder, GSAMDA offered a unique computer model for forecasting probable microbe–drug interactions.

We tested these techniques using their default settings and compared them using the fivefold CV experiment. AUC and AUPR values are used as indicators to evaluate the performance of NMGMDA, and the database we utilize is MDAD and aBioflm. The outcome was displayed in Table 3 and Fig. 4. Our suggested NMGMDA model has the greatest prediction performance of all the methods.

Methods	AUC		AUPR	
	MADA	aBioflm	MDAD	aBioflm
HMDAKATZ	0.90118 ± 0.00101	0.90023 ± 0.00212	0.23271 ± 0.00685	0.30669 ± 0.00771
LAGCN	0.86883 ± 0.00703	0.86413 ± 0.01091	0.35712 ± 0.00514	0.36715 ± 0.00556
HMDA-Pred	0.91776 ± 0.00297	0.91756 ± 0.00402	0.02361 ± 0.00093	0.02848 ± 0.00069
MNNMDA	0.96306 ± 0.00202	0.93156 ± 0.00223	0.18920 ± 0.00561	0.19621 ± 0.00782
GSAMDA	0.94917 ± 0.00053	0.93073 ± 0.01206	0.44363 ± 0.00072	0.45103 ± 0.00513
NMGMDA	0.99468 ± 0.00016	0.99467 ± 0.00017	0.90734 ± 0.00019	0.90330 ± 0.00018

Table 3. The AUCs and AUPRs of compared methods based on databases MDAD and aBioflm under fivefold CV.
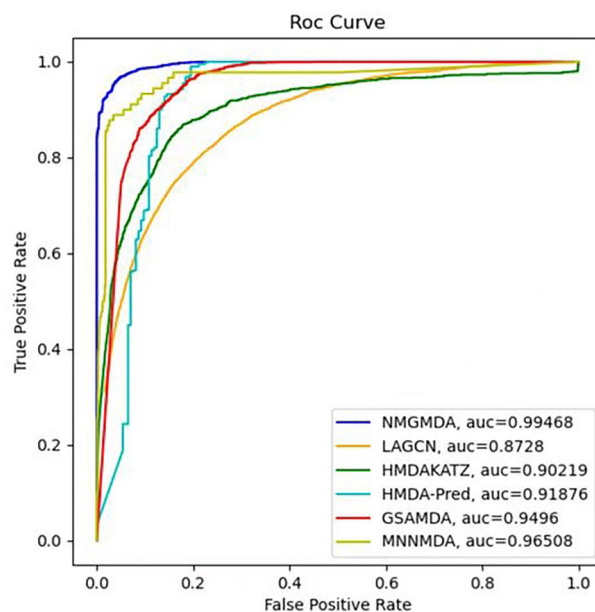
Figure 4. ROC curves based on the MDAD database for six competitive methods.

Case study

To test the NMGMDA model's real prediction power, we chose two well-known drugs—Ciprofloxacin and Moxifloxaxin—as well as two common microbes—Human immunodeficiency virus type 1 and Mycobacterium tuberculosis—for case studies.

Ciprofloxacin is an organic molecule with excellent bactericidal effect and broad-spectrum antibacterial activity[33]. It has shown to be a successful treatment for both acute and chronic urinary tract infections, as well as a variety of systemic infections[34]. Staphylococcus aureus[35], Haemophilus influenzae[36] and Stenotrophomonas maltophilia[37] are all susceptible to its antibacterial properties. Based on the predicted score, ranked the Ciprofloxacin-related microbes scores from highest to lowest, and chose the top 20 microbes for validation after deleting the 10 associations that are currently on MDAD database. As indicated in Table 4, 19 of the top 20 anticipated microbes connected to Ciprofloxacin have been verified by published research in PubMed. Moreover, Moxifloxacin[39] belongs to the quinolone drugs class, which mostly used to treat infections of the skin and soft tissues in adults as well as upper and lower respiratory tract infections[38,39]. According to the literature[40], Moxifloxacin is an effective treatment for Stenotrophomonas maltophilia keratitis. As indicated in Tables 5, after removing the 4 known associations on MDAD database, we discovered 17 microbes that had been verified by PubMed literature among the top 20 predicted microbes associated with moxifloxacin.

Regarding microbes, the first microbe is Human immunodeficiency virus type 1 (HIV-1), which is a virus capable of attacking the immune system in humans, and causes AIDS, an extremely dangerous infectious illness[41]. HIV-1 has been widely studied in relation to various medicines. Saquinavir, for instance, has been shown to be an effective treatment for HIV-1-infected individuals who have diarrhea and/or wasting syndrome by Hervé Trout[42]. According to literature[43], the first-line protease inhibitor that is generally suggested in the initial treatment regimen for people with HIV-1 infection is lopinavir/ritonavir. After removing the 26 known associations on MDAD

Microbe	Evidence	Microbe	Evidence
Enteric bacteria and other eubacteria	PMID: 36682905	*Acinetobacter baumannii*	PMID: 34098109
Micrococcus luteus	PMID: 3010848	*Klebsiella pneumoniae*	PMID: 29055688
Bacillus anthracis	PMID: 23569822	*Burkholderia pseudomallei*	PMID: 35972245
Helicobacter pylori	PMID: 24837413	*Candida albicans*	PMID: 35404123
Listeria monocytogenes	PMID: 34068252	*Vibrio anguillarum*	PMID: 36735199
Burkholderia cenocepacia	PMID: 25267676	*Actinomyces oris*	Unconfrmed
Burkholderia multivorans	PMID: 34524889	*Bacillus cereus*	PMID: 26358183
Mycobacterium avium	PMID: 30012773	*Vibrio vulnificus*	PMID: 28971862
Vibrio harveyi	PMID: 27247095	*Streptococcus mutans*	PMID: 33402618
Klebsiella planticola	PMID: 36452290	*Streptococcus epidermidis*	PMID: 34948159

Table 4. The top 20 Ciprofloxacin associated candidate microbes on MDAD. The top 10 microbes are listed in the first column, while the top 11–20 microbes are listed in the third column.

Microbe	Evidence	Microbe	Evidence
Staphylococcus aureus	PMID: 33936821	*Burkholderia multivorans*	Unconfrmed
Escherichia coli	PMID: 34653694	*Providencia stuartii*	Unconfrmed
Pseudomonas aeruginosa	PMID: 23662986	*Burkholderia cenocepacia*	PMID: 33120688
Bacillus subtilis	PMID: 30036828	*Mycobacterium tuberculosis*	PMID: 33951360
Staphylococcus epidermidis	PMID: 15718814	*Enterococcus faecalis*	PMID: 26349832
Streptococcus mutans	PMID: 29160117	*Vibrio harveyi*	Unconfrmed
Staphylococcus epidermis	PMID: 31516359	*Klebsiella planticola*	PMID: 32577260
Salmonella enterica	PMID: 34439014	*Listeria monocytogenes*	PMID: 18299415
Micrococcus luteus	PMID: 16152924	*Candida glabrata*	PMID: 20455400
Shigella flexneri	PMID: 28483960	*Streptococcus pneumoniae*	PMID: 22407042

Table 5. The top 20 Moxifoxacin associated candidate microbes on MDAD. The top 10 microbes are listed in the first column, while the top 11–20 microbes are listed in the third column.

database, we discovered 16 (see Table 6) drugs that had been validated by PubMed literatures among the top 20 anticipated microbes associated with Human immunodeficiency virus type 1. Mycobacterium tuberculosis is the second microbes used in the case study. Mycobacterium tuberculosis is the pathogen that causes tuberculosis[44], and many microbes, including ciprofloxacin[45] and triclosan[46], have been shown to be associated with it. After removing the 14 known associations on MDAD database, Table 7 indicates that of the top 20 candidate drugs, 14 were linked to Mycobacterium tuberculosis.

In conclusion, these two sets of case studies further demonstrate how the NMGMDA model may anticipate the association between microbes and drugs.

Drug	Evidence	Drug	Evidence
Delavirdine	PMID: 8807058	Ceftazidime	PMID: 7979856
Nelfinavir	PMID: 10223697	Amikacin	PMID: 16371246
LL-37	PMID: 24821067	Tobramycin	PMID: 27387258
Farnesol	PMID: 22677126	Cefixime	PMID: 9677171
Vancomycin	PMID: 32328074	Curcumin	PMID: 19345695
Epigallocatechin Gallate	PMID: 11684313	N-Acetylcysteine	PMID: 32796068
Toremifene	Unconfrmed	IDR-1018	Unconfrmed
Norspermidine	Unconfrmed	Metronidazole	PMID: 18444793
Hamamelitannin	PMID: 8693037	Terpinene-4-ol	Unconfrmed
Ciprofloxacin	PMID: 9566552	Carvacrol	PMID: 32461309

Table 6. The top 20 Human immunodeficiency virus type 1 associated candidate drugs on MDAD. The top 10 drugs are listed in the first column, while the top 11–20 drugs are listed in the third column.

Drug	Evidence	Drug	Evidence
Curcumin	PMID:27012592	Norspermidine	Unconfrmed
Epigallocatechin gallate	PMID:33463343	Moxifloxacin	PMID: 33951360
Metronidazole	PMID:18491971	Pleurocidin	Unconfrmed
BMAP-28	Unconfrmed	Esculetin	Unconfrmed
Vancomycin	PMID:33508482	Gentamicin	PMID: 36258995
Ceftazidime	PMID:11527042	Amikacin	PMID: 34314673
Cefixime	Unconfrmed	Silver nanoparticles	PMID: 32280217
Farnesol	PMID:16041726	Carboxymethyl chitosan	PMID: 30007629
Azithromycin	PMID: 32781595	Toremifene	Unconfrmed
LL-37	PMID: 26351280	Indole	PMID: 34383995

Table 7. The top 20 *Mycobacterium tuberculosis* associated candidate drugs on MDAD. The top 10 drugs are listed in the first column, while the top 11–20 drugs are listed in the third column.

Discussion and conclusion

The association between drugs and microbes has enormous significance for the treatment of diseases, according to biomedical studies. Consequently, a powerful computational prediction model could help researchers find additional microbe–drug associations and improve illness therapy.

By combining the NNM and GAT frameworks, we suggested a unique model in this study called NMGNAD to forecast potential microbe–drug associations. In NMGNAD, we first combined the drug similarity network, the known microbe–drug associations, and the similarity and association information between nodes to create a new microbe–drug heterogeneous network. The correlation scores between microbes and drugs were then predicted using the NNM model and the GAT model. In order to get the forecast results, we weighted average these two anticipated scores. According to experimental findings, NMGMAD outperformed state-of-the-art methods and produced acceptable case study outcomes.

Although NMGMDA can produce good prediction performance, there are still certain restrictions. First off, a few drug names in both databases are not accessible now, and the fact that they are no longer being updated will reduce the number of known connections that are available and have an impact on how the model is used in practice. Thus, we might think about creating a microbe–drug database that is more extensive. Then, to increase the precision of model predictions, we can think about adding more biological data to enhance the characteristics of drugs and microbes, such as data on drug side effects, data on the relationship between germs and diseases, and data on the association between drugs and diseases.

Data availability

The original contributions presented in the study are included in the article, further inquiries can be directed to the corresponding authors.

References

1. Huttenhower, C., Gevers, D., Knight, R., Badger, J. H., Human Microbiome Project Consortium. Structure, function and diversity of the healthy human microbiome. *Nature* **486**, 207–214. https://doi.org/10.1038/nature11234 (2012).
2. Liang, C., Changlu, Q., He, Z., Tongze, F. & Xue, Z. gutMDisorder: A comprehensive database for dysbiosis of the gut microbiota in disorders and interventions. *Nucleic Acids Res.* **48**, 554–560. https://doi.org/10.1093/nar/gkz843 (2020).
3. Gill, S. R. *et al.* Metagenomic analysis of the human distal gut microbiome. *Science* **312**, 1355–1359. https://doi.org/10.1126/science.1124234 (2006).
4. Ventura, M. *et al.* Genome-scale analyses of health-promoting bacteria: Probiogenomics. *Nat. Rev. Microbiol.* **7**, 61–71. https://doi.org/10.1038/nrmicro2047 (2009).
5. Sommer, F. & Bäckhed, F. The gut microbiota-masters of host development and physiology. *Nat. Rev. Microbiol.* **11**, 227–238. https://doi.org/10.1038/nrmicro2974 (2013).
6. Ley, R. E., Turnbaugh, P. J., Klein, S. & Gordon, J. I. Human gut microbes associated with obesity. *Nature* **444**, 1022–1023. https://doi.org/10.1038/4441022a (2006).
7. Durack, J. & Lynch, S. V. The gut microbiome: Relationships with disease and opportunities for therapy. *Rockefeller Univ. Press* **216**, 20–40. https://doi.org/10.1084/jem.20180448 (2019).
8. Schwabe, R. F. & Jobin, C. The microbiome and cancer. *Nat. Rev. Cancer* **13**, 800–812. https://doi.org/10.1038/nrc3610 (2013).
9. McCoubrey, L. E., Gaisford, S., Orlu, M. & Basit, A. W. Predicting drug-microbiome interactions with machine learning. *Biotechnol. Adv.* **54**, 107797. https://doi.org/10.1016/j.biotechadv.2021.107797 (2022).
10. Zimmermann, M., Zimmermann-Kogadeeva, M., Wegmann, R. & Goodman, A. L. Mapping human microbiome drug metabolism by gut bacteria and their genes. *Nature* **570**, 462–467. https://doi.org/10.1038/s41586-019-1291-3 (2019).
11. Wang, L., Tan, Y., Yang, X., Kuang, L. & Ping, P. Y. Review on predicting pairwise relationships between human microbes, drugs and diseases: From biological data to computational models. *Brief. Bioinform.* **23**, 1–25. https://doi.org/10.1093/bib/bbac080 (2022).
12. Sun, Y. Z. *et al.* MDAD: A special resource for microbe–drug associations. *Front. Cell. Infect. Microbiol.* **8**, 424. https://doi.org/10.3389/fcimb.2018.00424 (2018).
13. Akanksha, R., Anamika, T., Shivangi, S. & Manoj, K. aBiofilm: A resource of anti-biofilm agents and their potential implications in targeting antibiotic drug resistance. *Nucleic Acids Res.* **46**, 894–900. https://doi.org/10.1093/nar/gkx1157 (2018).
14. Zhu, L., Duan, G., Yan, C. & Wang, J . Prediction of microbe–drug associations based on KATZ measure. In *2019 IEEE International Conference on Bioinformatics and Biomedicine*. https://doi.org/10.1109/BIBM47256.2019.8983209 (2019).
15. Long, Y. & Luo, J. Association mining to identify microbe drug interactions based on heterogeneous network embedding representation. *IEEE J. Biomed. Health Inform.* **25**, 266–275. https://doi.org/10.1109/jbhi.2020.2998906 (2021).
16. Ma, Y. & Liu, Q. Generalized matrix factorization based on weighted hypergraph learning for microbe–drug association prediction. *Comput. Biol. Med.* **145**, 105503. https://doi.org/10.1016/j.compbiomed.2022.105503 (2022).
17. Yang, H., Ding, Y., Tang, J. & Gao, F. Inferring human microbe–drug associations via multiple kernel fusion on graph neural network. *Knowl. Based Syst.* **238**, 107888. https://doi.org/10.1016/j.knosys.2021.107888 (2022).
18. Zhu, B. *et al.* NNAN: Nearest neighbor attention network to predict drug–microbe associations. *Front. Microbiol.* **13**, 846915. https://doi.org/10.3389/fmicb.2022.846915 (2022).
19. Tian, Z., Yu, Y., Fang, H., Xie, W. & Guo, M. Predicting microbe–drug associations with structure-enhanced contrastive learning and self-paced negative sampling strategy. *Brief. Bioinform.* **24**, 634. https://doi.org/10.1093/bib/bbac634 (2023).
20. Tan, Y. *et al.* GSAMDA: A computational model for predicting potential microbe–drug associations based on graph attention network and sparse autoencoder. *BMC Bioinform.* **23**, 492. https://doi.org/10.1186/s12859-022-05053-7 (2022).
21. Yang, H., Ding, Y., Tang, J. & Guo, F. Inferring human microbe–drug associations via multiple kernel fusion on graph neural network. *Knowl.-Based Syst.* **28**, 107888. https://doi.org/10.1016/j.knosys.2021.107888 (2022).
22. Ma, Q., Tan, Y. & Wang, L. GACNNMDA: A computational model for predicting potential human microbe–drug associations based on graph attention network and CNN-based classifier. *BMC Bioinform.* **24**, 35. https://doi.org/10.1186/s12859-023-05158-7 (2023).
23. Yoo, S., Wang, Z. & Seo, J. Adaptive weighted nuclear norm minimization for removing speckle noise from optical coherence tomography images. In *2019 41st Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*. https://doi.org/10.1109/embc.2019.8857208 (2019).
24. Veličković, P., *et al.* Graph attention networks. arXiv:1710.10903. https://doi.org/10.48550/arXiv.1710.10903 (2017).

25. Szklarczyk, D. *et al.* STRING v11: Protein–protein association networks with increased coverage, supporting functional discovery in genome-wide experimental datasets. *Nucleic Acids Res.* **47**, 607–613. https://doi.org/10.1093/nar/gky1131 (2019).
26. Kamneva, O. K. Genome composition and phylogeny of microbes predict their co-occurrence in the environment. *PLoS Comput. Biol.* **13**, 1005366. https://doi.org/10.1371/journal.pcbi.1005366 (2017).
27. Hattori, M., Tanaka, N., Kanehisa, M. & Goto, S. SIMCOMP/SUBCOMP: Chemical structure search servers for network analyses. *Nucleic Acids Res.* **38**, 652–656. https://doi.org/10.1093/nar/gkq367 (2010).
28. Yang, M., Luo, H., Li, Y. & Wang, J. Drug repositioning based on bounded nuclear norm regularization. *Bioinformatics* **35**, 455–463. https://doi.org/10.1093/bioinformatics/btz331 (2019).
29. Candès, E. & Recht, B. Simple bounds for recovering low-complexity models. *Math. Program.* **141**, 577–589. https://doi.org/10.1007/s10107-012-0540-0 (2012).
30. Fan, Y., Chen, M., Zhu, Q. & Wang, W. Inferring disease-associated microbes based on multi-data integration and network consistency projection. *Front. Bioeng. Biotechnol.* **8**, 831. https://doi.org/10.3389/fbioe.2020.00831 (2020).
31. Yu, Z., Huang, F., Zhao, X., Xiao, W. & Zhang, W. Predicting drug-disease associations through layer attention graph convolutional network. *Brief. Bioinform.* **22**, bbaa243. https://doi.org/10.1093/bib/bbaa243 (2021).
32. Haiyan, L. *et al.* MNNMDA: Predicting human microbe–disease association via a method to minimize matrix nuclear norm. *Comput. Struct. Biotechnol. J.* **21**, 1414–1423. https://doi.org/10.1016/j.csbj.2022.12.053 (2023).
33. Terp, D. K. & Rybak, M. J. Ciprofoxacin. *Drug Intell. Clin. Pharm.* **35**, 373–447. https://doi.org/10.1177/1060028087021007-801 (1988).
34. Campoli-Richards, D. M. *et al.* Ciprofloxacin. A review of its antibacterial activity, pharmacokinetic properties and therapeutic use. *J. Drugs* **35**, 373–447. https://doi.org/10.2165/00003495-198835040-00003 (1988).
35. Warraich, A. A. *et al.* Evaluation of anti-biofilm activity of acidic amino acids and synergy with ciprofloxacin on *Staphylococcus aureus* biofilms. *J. Sci. Rep.* **10**, 9021. https://doi.org/10.1038/s41598-020-66082-x (2020).
36. Kosikowska, U., Andrzejczuk, S., Plech, T. & Malm, A. Inhibitory effect of 1,2,4-triazole-ciprofloxacin hybrids on *Haemophilus parainfluenzae* and Haemophilus influenzae biofilm formation invitro under stationary conditions. *Res. Microbiol.* **167**, 647–654. https://doi.org/10.1016/j.resmic.2016.05.009 (2016).
37. Ba, B. B. *et al.* Activities of ciprofloxacin and moxifloxacin against *Stenotrophomonas maltophilia* and emergence of resistant mutants in an in vitro pharmacokinetic-pharmacodynamic model. *Antimicrob. Agents Chemother.* **48**, 946–953. https://doi.org/10.1128/aac.48.3.946-953.2004 (2004).
38. Balfour, J. A. & Wiseman, L. R. Moxifloxacin. *Drugs* **59**, 115–139. https://doi.org/10.2165/00003495-199957030-00007 (1999).
39. Jacobsen, F. *et al.* Efficacy of topically delivered moxifloxacin against wound infection by *Pseudomonas aeruginosa* and methicillin-resistant *Staphylococcus aureus*. *Antimicrob. Agents Chemother.* **55**, 2325–2334. https://doi.org/10.1128/aac.01071-10 (2011).
40. Felix, G. *et al.* Efficacy of antibiotic treatment of implant-associated *Staphylococcus aureus* infections with moxifloxacin, flucloxacillin, rifampin, and combination therapy: An animal study. *Drug Design Dev. Ther.* **11**, 1729–1736. https://doi.org/10.2147/dddt.s138888 (2017).
41. Spector, S. A. Human immunodefciency virus type-1. *Ref. Module Biomed. Sci.* **11**, 1–12. https://doi.org/10.1016/B978-0-12-801238-3.00088-X (2014).
42. Hervé, T. *et al.* Enhanced saquinavir exposure in human immunodeficiency virus Type 1-infected patients with diarrhea and/or wasting syndrome. *Antimicrob. Agents Chemother.* **48**, 538–545. https://doi.org/10.1128/aac.48.2.538-545.2004 (2004).
43. Kaplan, S. S. & Hicks, C. B. Safety and antiviral activity of lopinavir/ritonavir-based therapy in human immunodeficiency virus type 1 (HIV-1) infection. *J. Antimicrob. Chemother.* **56**, 273–276. https://doi.org/10.1093/jac/dki209 (2005).
44. Koch, A. & Mizrahi, V. Mycobacterium tuberculosis. *Trends Microbiol.* **26**, 555–556. https://doi.org/10.1016/j.tim.2018.02.012 (2018).
45. Chen, C. H., Shih, J. F., Lindholm-Levy, P. J. & Heifets, L. B. Minimal inhibitory concentrations of rifabutin, ciprofloxacin, and ofloxacin against *Mycobacterium tuberculosis* isolated before treatment of patients in Taiwan. *Am. Rev. Respir. Dis.* **140**, 987–989. https://doi.org/10.1164/ajrccm/140.4.987 (1989).
46. Parikh, S. L., Xiao, G. & Tonge, P. J. Inhibition of InhA, the enoyl reductase from *Mycobacterium tuberculosis*, by triclosan and isoniazid. *Biochemistry* **39**, 7645–7650. https://doi.org/10.1021/bi0008940 (2000).

Acknowledgements

Author contributions

M.L., X.L. and L.W. produced the main ideas, and did the modeling, computation and analysis and also wrote the manuscript. Q.C., L.W. and B.Z. provided supervision and effective scientific advice and related ideas, research design guidance, and added value to the article through editing and contributing completions. All authors contributed to the article and approved the submitted version.

Funding

Competing interests

The authors declare no competing interests.

Additional information

Correspondence and requests for materials should be addressed to Q.C., B.Z. or L.W.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.