

## Article

# Biology-Informed Matrix Factorization: An AI-Driven Framework for Enhanced Drug Repositioning

Yangyang Wang <sup>1,†</sup>, Yaping Wang <sup>2,†</sup>, Ya Hu <sup>3</sup> and Jihan Wang <sup>2,\*</sup><sup>1</sup> School of Physics and Electronic Information, Yan'an University, Yan'an 716000, China<sup>2</sup> Yan'an Medical College, Yan'an University, Yan'an 716000, China<sup>3</sup> Department of Medical College, Hunan Polytechnic of Environment and Biology, Hengyang 421000, China\* Correspondence: [jihanwang@yau.edu.cn](mailto:jihanwang@yau.edu.cn)

† These authors contributed equally to this work.

**Simple Summary:** Finding new medical uses for existing drugs can help patients faster and at a lower cost compared to developing entirely new drugs. This process, called drug repositioning, uses the information we already know about approved drugs to see if they might treat other diseases. In this study, we created a computer model that combines drug and disease similarities based on biological knowledge, making predictions about which drugs might be effective for diseases they are not currently used to treat. Our method uses a mathematical technique called matrix factorization, enhanced by biological information, to improve accuracy and reliability. We tested the model using well-known data and showed that it works better than existing approaches. We also looked at several real drugs and found new disease connections for each one, many of which were supported by trusted medical databases. This approach helps researchers discover new treatment options faster and more efficiently, which could benefit public health and reduce costs in drug development.

**Abstract:** Advances in artificial intelligence (AI) and intelligent computing have significantly accelerated drug discovery by enabling accurate modeling of complex biomedical relationships. Among these efforts, drug repositioning—identifying novel therapeutic uses for approved or investigational drugs—offers a cost-effective and time-efficient alternative to de novo drug development. While non-negative matrix factorization (NMF) has been widely adopted for uncovering latent drug–disease associations, conventional implementations often neglect the biological context that underpins these relationships. In this work, we propose a novel NMF-based drug repositioning model that incorporates biological context (NMFIBC), which integrates drug and disease similarity networks through graph-regularized optimization to enhance predictive performance. This design enhances both the robustness and interpretability of association prediction. Extensive benchmarking on multiple gold-standard datasets demonstrates that NMFIBC outperforms existing methods across a range of metrics, including AUC, precision, and F1-score. Moreover, case studies involving clinically relevant drugs validate the biological plausibility of the predicted associations using public databases such as DrugBank, CTD, and KEGG. The proposed framework provides a powerful, context-aware AI strategy for discovering actionable insights in drug repositioning research.

**Keywords:** artificial intelligence; drug repositioning; non-negative matrix factorization; biological context; graph regularization; drug–disease association prediction



Academic Editors: Yong Chen and Milana Frenkel-Morgenstern

Received: 18 April 2025

Revised: 8 May 2025

Accepted: 14 May 2025

Published: 15 May 2025

**Citation:** Wang, Y.; Wang, Y.; Hu, Y.; Wang, J. Biology-Informed Matrix Factorization: An AI-Driven Framework for Enhanced Drug Repositioning. *Biology* **2025**, *14*, 549. <https://doi.org/10.3390/biology14050549>

**Copyright:** © 2025 by the authors. Licensee MDPI, Basel, Switzerland.

This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

## 1. Introduction

Drug repositioning, also known as drug repurposing, is the process of identifying new therapeutic uses for existing drugs beyond their original medical indications [1,2]. This strategy leverages established pharmacokinetic, pharmacodynamic, and toxicity profiles to accelerate development, reduce costs, and lower the risk of failure in clinical trials. Repositioning can involve both approved drugs and those discontinued for reasons unrelated to safety. The pharmaceutical industry continues to face substantial challenges in traditional drug discovery, including high costs, extended timelines, and high attrition rates [2]. Drug repositioning offers a practical alternative by reassessing existing compounds for new indications [3]. Compared to de novo drug development, which may take over a decade and cost upwards of \$2 billion, repositioning significantly reduces both development time and financial burden. Moreover, as repositioned drugs have already undergone preclinical and clinical evaluations, they carry a lower risk of safety-related failure [4,5].

In recent years, a variety of computational approaches have been proposed to facilitate drug repositioning, including signature matching, network-based models, and matrix decomposition-based techniques [4]. Signature matching compares gene expression profiles of diseases and drugs to identify therapeutic candidates that may reverse disease phenotypes. Network-based methods explore biological interaction networks (e.g., protein–protein interaction, gene regulation) to uncover latent links between drugs and diseases. Matrix decomposition methods—such as non-negative matrix factorization (NMF) [6–8] and singular value decomposition (SVD) [9,10]—aim to extract low-rank representations from drug–disease association matrices, revealing latent associations that are often imperceptible in the high-dimensional space [11].

Among the above approaches, NMF has attracted significant attention due to its scalability, interpretability, and effectiveness in identifying hidden structures. For example, Zhang et al. proposed feature-derived graph regularized matrix factorization (FGRMF) [12], which predicts drug side effects by incorporating drug similarity features. The similarity-constrained matrix factorization method (SCMFDD) [13] extended this approach by adding regularization terms for both drug and disease similarity. Sadeghi et al. introduced NMFDR [6], a network-based NMF variant for predicting new drug–disease associations. Other recent developments include bounded nuclear norm regularization (BNNR) [14], which mitigates cold start problems, and hybrid models that combine matrix factorization with semantic diffusion frameworks [15]. However, existing NMF-based methods often neglect the biological context underlying drug–disease associations, focusing primarily on numerical decompositions without integrating domain-specific knowledge. This omission may limit both prediction accuracy and the interpretability of results.

To address these limitations, we propose a novel non-negative matrix factorization-based drug repositioning model that incorporates biological context (NM-FIBC). Our framework embeds biological similarity networks—capturing functional or semantic relationships among drugs and diseases—into a graph-regularized NMF optimization scheme. This design enhances predictive accuracy and ensures that inferred associations are not only statistically robust but also biologically meaningful.

## 2. Materials and Methods

### 2.1. Datasets

We employed two gold-standard datasets (Cdataset [16] and Fdataset [17]) to evaluate the performance of the proposed method in predicting novel drug–disease associations. These datasets vary in size, containing different numbers of drugs, diseases, and known associations, as detailed in Table 1. Each dataset includes precomputed drug similarity

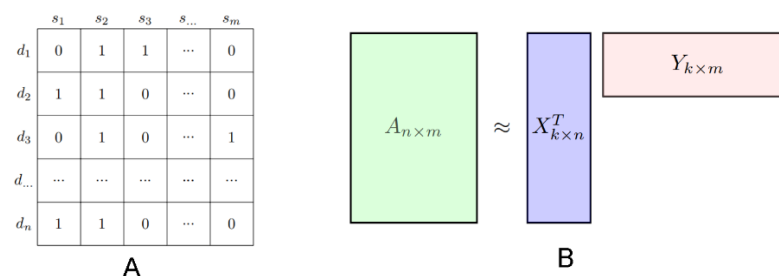
matrices, disease similarity matrices, and a binary association matrix that encodes known drug–disease interactions.

**Table 1.** Summary of drug–disease similarity matrices and known associations in the benchmark datasets.

Dataset	Drug Similarity Matrix	Disease Similarity Matrix	Known Associations
Cdataset	$663 \times 663$	$409 \times 409$	$409 \times 663$
Fdataset	$593 \times 593$	$313 \times 313$	$313 \times 593$

## 2.2. Standard NMF

Suppose we have a set of  $n$  drugs and  $m$  disease, denoted as  $D = \{d_1, d_2, \dots, d_n\}$  and  $S = \{s_1, s_2, \dots, s_m\}$ , respectively. The known side effects of these drugs, that is, the drug–side effect associations, are naturally represented by an  $n \times m$  matrix of  $A$ . As illustrated in Figure 1A, the binary value “0” or “1” of entry  $A_{i,j}$  indicates the absence or presence of disease  $s_j$  for drug  $d_i$ . Matrix factorization plays a pivotal role in various applications, particularly in recommendation systems, data analysis, and machine learning. Its primary function is to decompose a large, complex matrix into a product of two or more smaller, lower-rank matrices, as shown in Figure 1B. The matrix of  $A$  can be decomposed into two low-rank matrices  $X$  and  $Y$ , where  $A \approx X^T Y$ ,  $X \in R^{k \times n}$ ,  $Y \in R^{k \times m}$  and  $\text{rank}(X) = \text{rank}(Y)$ . Generally, the selection of the parameter may affect the accuracy of the final drug repositioning. From Figure 1, it is evident that the drug repurposing problem can be addressed through matrix factorization. This approach leverages the inherent structure of the data to decompose the complex relationships between drugs and their potential indications into more manageable components, thereby facilitating the discovery of new uses for existing drugs.



**Figure 1.** Overview of the standard non-negative matrix factorization (NMF) workflow for drug–disease association prediction. (A) Binary drug–disease association matrix representing known interactions. (B) Decomposition of the matrix into two lower-dimensional matrices capturing latent drug and disease features for use in predictive modeling.

NMF is a valuable tool in data science due to its ability to provide insights into the composition of non-negative datasets and its interpretability of the underlying structure. The workflow of NMF is shown in Figure 1B. It is used to decompose a non-negative data matrix  $A$  into two or more non-negative matrices  $X$  and  $Y$ , whose product approximates the original matrix, and can be formulated by Equation (1):

$$\min_{W \geq 0, H \geq 0} \|A - X^T Y\|_F^2, \text{ s.t. } X \geq 0, Y \geq 0 \quad (1)$$

where  $\|\cdot\|_F$  is the Frobenius norm,  $X \in R^{k \times n}$  and  $Y \in R^{k \times m}$  ( $k \ll \min\{m, n\}$ ) are the latent matrix,  $X_i$  and  $Y_j$  can be regarded as drug-specific and disease-specific latent feature vectors, and  $A_{ij} = X_i^T Y_j$ . There are several ways in which  $X$  and  $Y$  can be obtained, and the most popular method implemented by Lee et al. can be found in [18].



potential candidates by the NMFIBC model may offer a foundation for deeper analytical investigation and subsequent empirical verification. Based on the drug similarity constraints and Equation (2), the objective function of NMFIBC can be formulated as Equation (3):

$$\begin{aligned} L = & \min \|A - X^T Y\|_F^2 + \mu(\|X\|_F^2 + \|Y\|_F^2) + \lambda(\|X^T X - S_d\|_F^2 + \|Y^T Y - S_e\|_F^2) \\ = & \sum_{ij} (A_{ij} - x_i^T y_j)^2 + \mu \left( \sum_i \|x_i\|^2 + \sum_j \|y_j\|^2 \right) \\ & + \lambda \left( \sum_i \|x_i^T x_i - S_d\|^2 + \sum_j \|y_j^T y_j - S_e\|^2 \right) \\ \text{s.t. } & X \geq 0, Y \geq 0 \end{aligned} \quad (3)$$

## 2.4. Optimization Algorithm

In this section, we use gradient descent to derive the solutions for the two latent feature matrices,  $X$  and  $Y$ , based on the objective function in Equation (3). In order to solve the optimization problem of Equation (3), we introduce Lagrange multipliers,  $\Phi = [\varphi_{ik}]$  and  $\Psi = [\psi_{jk}]$ , to implement the constraints on  $X \geq 0, Y \geq 0$  [23]. The objective function can be transformed into Equation (4).

$$\begin{aligned} L_f = & \text{Tr}(AA^T - AY^T X - X^T Y A^T + X^T Y Y^T X) + \mu \text{Tr}(XX^T + YY^T) \\ & + \lambda \text{Tr}(S_d S_d^T - S_d X^T X - X^T X S_d + X^T X X^T X) \\ & + \lambda \text{Tr}(S_e S_e^T - S_e Y^T Y - Y^T Y S_e + Y^T Y Y^T Y) \\ & + \text{Tr}(\Phi X^T) + \text{Tr}(\Phi Y^T) \end{aligned} \quad (4)$$

The partial derivatives with respect to  $X$  and  $Y$  are as Equations (5) and (6):

$$\frac{\partial L}{\partial X} = -Y A^T + Y Y^T X + \mu X + 2\lambda(-X(S_d) + X X^T X) + \Phi \quad (5)$$

$$\frac{\partial L}{\partial Y} = -X A + X X^T Y + \mu Y + 2\lambda(-Y(S_e) + Y Y^T Y) + \Psi \quad (6)$$

By using the Karush–Kuhn–Tucker (KKT) conditions [24], Equations (5) and (6) can be transformed into Equations (7) and (8):

$$\left(-Y A^T + Y Y^T X\right)_{ik} X_{ik} + (\mu X)_{ik} X_{ik} + \left(2\lambda(-X S_d + X X^T X)\right)_{ik} X_{ik} = 0 \quad (7)$$

$$\left(-X A + X X^T Y\right)_{jk} Y_{jk} + (\mu Y)_{jk} Y_{jk} + \left(2\lambda(-Y S_e + Y Y^T Y)\right)_{jk} Y_{jk} = 0 = 0 \quad (8)$$

Thus, the updating rules for  $X$  and  $Y$  can be obtained, as shown in Formulas (9) and (10):

$$X_{ik} \leftarrow X_{ik} \frac{(Y A^T + 2\lambda X S_d)_{ik}}{(Y Y^T X + \mu X + 2\lambda X X^T X)_{ik}} \quad (9)$$

$$Y_{jk} \leftarrow Y_{jk} \frac{(X A + 2\lambda Y S_e)_{jk}}{(X X^T Y + \mu Y + 2\lambda Y Y^T Y)_{jk}} \quad (10)$$

Upon adequate convergence of  $X$  and  $Y$ , or upon reaching a predefined number of iterations, the predictive matrix  $\hat{A}$  can be accurately reconstructed as the product of obtained  $X$  and  $Y$ ,  $\hat{A} \approx X^T Y$ .

Algorithm 1 summarizes the procedure of NMFIBC for drug–disease association prediction.

**Algorithm 1:** Optimization algorithm for NMFIBC

---

**Input:** Drug similarity matrix,  $X \in R^{n \times n}$ ;  
Disease similarity matrix,  $Y \in R^{m \times m}$ ;  
Drug – side effect association matrix,  $A \in R^{n \times m}$ ;  
The latent dimension of feature space,  $k < \min\{m, n\}$ ;  
Regularization parameter,  $\mu > 0$ ,  $\lambda > 0$ ;  
Maximum iterations  $M$ .  
**Output :** The prediction matrix  $\hat{A}$ ;  
Initialize  $X \in R^{n \times n}$  and  $Y \in R^{m \times m}$ ;  
 $times \leftarrow 0$ ;  
**while**  $times < M$  **do**  
    update  $X$  by using Equation (9);  
    update  $Y$  by using Equation (10);  
     $times \leftarrow times + 1$ ;  
**end**  
 $\hat{A} \leftarrow XY^T$ ;

---

**2.5. Evaluation Metrics**

In this section, we present a comprehensive evaluation of the performance of five distinct algorithms (IMCMDA [25], NCPMDA [26], RLSMDA [27], and SIMCLDA [28]) on two datasets, Cdataset and Fdataset. The metrics of interest include Area Under the Curve (AUC), Area Under the Precision-Recall Curve (AUPR), Accuracy (Acc), Sensitivity or Recall (Sen), Specificity (Spe), Precision (Pre), and F1 Score (F1), as shown in Equations (11)–(16). These metrics provide a holistic view of the classification capabilities of each algorithm. All the experiments were conducted using MATLAB 2023b on Windows 10, running on an Intel(R) Core (TM) i5-12400F at 2.50 GHz.

$$SN = \frac{TP}{TP + FN} \quad (11)$$

$$SP = \frac{TN}{TN + FP} \quad (12)$$

$$Acc = \frac{TP + TN}{TP + TN + FP + FN} \quad (13)$$

$$precision = \frac{TP}{TP + FP} \quad (14)$$

$$recall = \frac{TP}{TP + FN} \quad (15)$$

$$F1 = \frac{2 * precision * recall}{precision + recall} \quad (16)$$

**3. Results****3.1. Performance Evaluation and Metric Analysis**

To assess the predictive capability of the proposed NMFIBC model, we conducted a comparative analysis against four state-of-the-art algorithms (IMCMDA, NCPMDA, RLSMDA, and SIMCLDA) on two benchmark datasets: Cdataset and Fdataset. The evaluation metrics include AUC, accuracy, sensitivity, specificity, precision, and F1 score. As summarized in Tables 2 and 3, NMFIBC achieves the highest AUC scores on both datasets (0.921 on Cdataset and 0.894 on Fdataset), outperforming all competing methods.



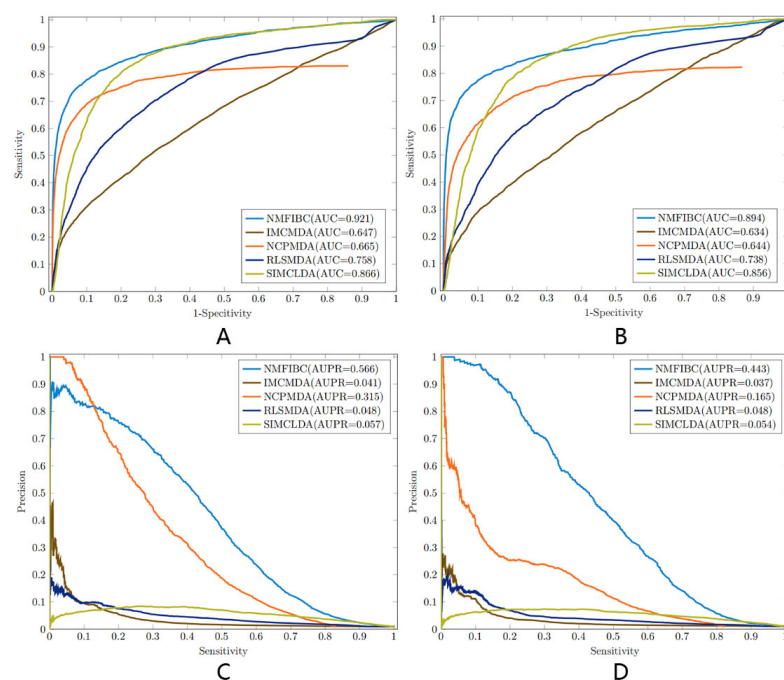
**Table 2.** Performance comparison of competing methods on Cdataset across multiple evaluation metrics.

	AUC	AUPR	Acc	Sen (Recall)	Spe	Pre	F1
NMFIBC	0.921	0.566	0.993	0.504	0.997	0.633	0.561
IMCMDA	0.647	0.041	0.979	0.129	0.988	0.090	0.106
NCPMDA	0.665	0.315	0.989	0.333	0.995	0.395	0.362
RLSMDA	0.758	0.048	0.979	0.149	0.987	0.096	0.116
SIMCLDA	0.866	0.057	0.954	0.392	0.959	0.083	0.136

**Table 3.** Performance comparison of competing methods on Fdataset across multiple evaluation metrics.

	AUC	AUPR	Acc	Sen (Recall)	Spe	Pre	F1
NMFIBC	0.894	0.443	0.990	0.430	0.996	0.501	0.463
IMCMDA	0.634	0.037	0.982	0.100	0.992	0.111	0.105
NCPMDA	0.644	0.165	0.981	0.343	0.988	0.224	0.271
RLSMDA	0.738	0.048	0.982	0.121	0.991	0.121	0.121
SIMCLDA	0.856	0.054	0.941	0.408	0.946	0.074	0.125

Beyond AUC, NMFIBC also demonstrates superior performance in accuracy (0.993 on Cdataset and 0.990 on Fdataset), along with consistently high precision and F1 scores. Notably, the model maintains a balanced trade-off between sensitivity and specificity, with specificity values exceeding 0.99, indicating robustness in distinguishing both positive and negative associations. In contrast, other models such as IMCMDA and RLSMDA show moderate accuracy but comparatively lower recall and F1 scores, suggesting potential overfitting or class imbalance issues. Figure 3 presents a visual comparison of AUC and AUPR scores across the five models, while Figure 4 illustrates their F1 scores. These results suggest that incorporating biological similarity networks through graph-regularized optimization enables NMFIBC to generalize more effectively, especially when dealing with sparse and noisy biomedical data.



**Figure 3.** AUC and AUPR scores of the five models (NMFIBC, IMCMDA, NCPMDA, RLSMDA, and SIMCLDA) evaluated on Cdataset and Fdataset. (A,B) AUC for Cdataset and Fdataset, respectively. (C,D) AUPR for Cdataset and Fdataset, respectively.

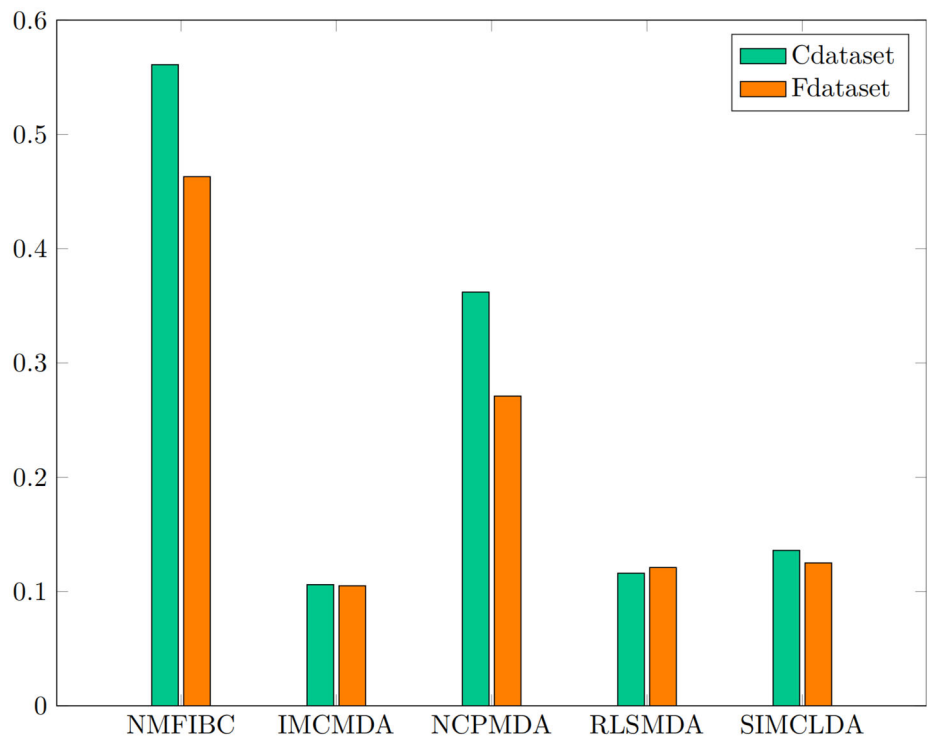


Figure 4. F1 Score of the five models on Cdataset and Fdataset.

3.2. Case Study Overview

To further assess the practical utility of NMFIBC, we conducted case studies on five representative drugs: Levodopa, Doxorubicin, Amantadine, Flecainide, and Tacrolimus. These drugs were selected due to their diverse clinical applications and well-documented pharmacological profiles. For each drug, we predicted the top five disease candidates that had no known associations in the original dataset. These predictions were ranked by their model-generated confidence scores and validated using biomedical databases such as Drug-Bank, Comparative Toxicogenomics Database (CTD), and Kyoto Encyclopedia of Genes and Genomes (KEGG). A summary of these case studies is provided in Table 4 (based on Cdataset) and Table 5 (based on Fdataset), which include the following: the original known associations (if any), the top five predicted diseases with scores, and supporting evidence from public databases or literature. These predictions serve as compelling demonstrations of how NMFIBC can uncover novel drug indications. Further interpretation and biological relevance of these results are provided in the Discussion Section.

Table 4. Top five predicted diseases for selected drugs based on Cdataset and their validation using public databases.

Drugs	Diseases (Existing Relations in Original Matrix)	Top Five Predicted Candidate Diseases (No Relation in Original Matrix)	Weight	Evidence
Levodopa (DB01235)	Paralysis agitans (168100)	Dementia (125320)	0.761	DB/KEGG
	Parkinson disease (168600)	Alzheimer disease 9 (608907)	0.571	DB/KEGG
	Parkinson disease 2 (600116)	Alzheimer disease (605055)	0.568	DB/KEGG
	Parkinson disease 7 (606324)	Alzheimer disease 2 (104310)	0.560	DB/KEGG
	Parkinson disease 15 (260300)	Alzheimer disease 5 (602096)	0.536	DB/KEGG
Doxorubicin (DB00997)	Mismatch repair cancer syndrome 1 (276300)	Renal cell carcinoma (144700)	0.734	DB/KEGG
	Breast cancer (114480)	Testicular germ cell tumor (273300)	0.692	DB
	Lymphoblastic leukemia (247640)	Small cell cancer of the lung (182280)	0.654	DB
	Leukemia (601626)	Leukemia (246470)	0.651	KEGG
	Lymphoma (236000)	Dohle bodies and leukemia (223350)	0.649	KEGG



Table 4. Cont.

Drugs	Diseases (Existing Relations in Original Matrix)	Top Five Predicted Candidate Diseases (No Relation in Original Matrix)	Weight	Evidence
Amantadine (DB00915)	Paralysis agitans (168100) Multiple sclerosis (126200) Popliteal pterygium syndrome (119500)	Parkinson's disease 7 (606324)	0.337	DB/KEGG/CTD
		Parkinson's disease 15 (260300)	0.325	DB/KEGG/CTD
		Schizophrenia (181500)	0.322	DB/KEGG
		Parkinson's disease (168600)	0.318	DB/KEGG/CTD
		Parkinson's disease 2 (600116)	0.318	DB/KEGG/CTD
Flecainide (DB01195)	Atrial fibrillation (607554)	Hypertension (608622)	0.688	[29]
		Renal failure (161900)	0.672	[29]
		Insensitivity to pain with hyperplastic Myelinopathy (147530)	0.520	Unknown
		Raynaud disease (179600)	0.413	Unknown
		Atrial fibrillation (608583)	0.404	DB/KEGG/CTD
Tacrolimus (DB00864)	Dermatitis (603165) Dermatitis (605805) Dermatitis (605804) Dermatitis (605844)	Allergic rhinitis (607154)	0.625	[30]
		Asthma (208550)	0.462	[31]
		Asthma (600807)	0.438	[31]
		Breast cancer (114480)	0.424	[32]
		Renal failure (161900)	0.396	[33]

Table 5. Top five predicted diseases for selected drugs based on Fdataset and their validation using public databases.

Drugs	Diseases (Existing Relations in Original Matrix)	Top Five Predicted Candidate Diseases (No Relation in Original Matrix)	Weight	Evidence
Levodopa (DB01235)	Paralysis agitans (168100) Restless legs syndrome (102300)	Parkinson's disease (168600)	0.548	DB/KEGG/CTD
		Insensitivity to pain with hyperplastic Myelinopathy (147530)	0.531	Unknown
		Dementia (125320)	0.451	DB/KEGG, [34]
		Renal failure (161900)	0.422	Unknown
		Attention deficit hyperactivity disorder (143465)	0.382	Unknown
Doxorubicin (DB00997)	Myeloma (254500) Breast cancer (114480) Neuroblastoma (256700) Leukemia (601626) Lymphoma (236000)	Small cell cancer of the lung (182280)	0.577	[35]
		Colorectal cancer (114500)	0.573	[36]
		Testicular germ cell tumor (273300)	0.530	[37]
		Kaposi sarcoma (148000)	0.518	DB/KEGG
		Esophageal cancer (133239)	0.513	[38]
Amantadine (DB00915)	Paralysis agitans (168100) Multiple sclerosis (126200) Popliteal pterygium syndrome (119500)	Dementia (125320)	0.365	DB/KEGG/CTD
		Parkinson's disease (168600)	0.363	DB/KEGG/CTD
		Restless legs syndrome (102300)	0.295	[39]
		Alzheimer's disease (104300)	0.227	DB/KEGG/CTD
		Alzheimer disease (605055)	0.216	DB/KEGG/CTD
Flecainide (DB01195)	Atrial fibrillation (607554)	Hypertension (608622)	0.597	[29]
		Renal failure (161900)	0.560	[29]
		Atrial fibrillation (608583)	0.524	DB/CTD, [29]
		Insensitivity to pain with hyperplastic Myelinopathy (147530)	0.463	Unknown
		Stroke (601367)	0.335	[40]
Tacrolimus (DB00864)	Dermatitis (603165)	Renal failure (161900)	0.582	[33]
		Hypertension (608622)	0.490	[41]
		Asthma (208550)	0.381	[31]
		Insensitivity to pain with hyperplastic Myelinopathy (147530)	0.376	Unknown
		Hypoparathyroidism (146255)	0.374	Unknown

#### 4. Discussion

The increasing complexity of biological systems and the explosion of high-dimensional biomedical data have placed AI at the forefront of modern drug discovery and systems biology. In this context, the integration of AI algorithms with domain-specific biological knowledge has become essential for deriving meaningful insights from heterogeneous biomedical datasets. Our proposed model, a nonnegative matrix factorization-based drug repositioning framework that incorporates biological context (NMFIBC), is an example of

such an approach. By embedding biological similarity networks into the matrix factorization process, NMFIBC addresses the limitations of traditional models and enhances the discovery of novel, biologically plausible drug–disease associations.

In these exploratory analyses, the NMFIBC algorithm was utilized to forecast novel therapeutic applications for existing medications in practical scenarios. During the discovery phase of fresh connections between pharmaceuticals and medical conditions, the existing links within a benchmark dataset served as our training material, while the uncharted pairs were designated as the pool of prospective associations. Upon employing the NMFIBC model to compute the predictive scores for the entire spectrum of potential drug–disease pairs, we ranked the candidate illnesses in a hierarchy based on the calculated scores specific to each medication. To ascertain the veracity of these forecasts, we hand-picked Levodopa, Doxorubicin, Amantadine, Flecainide, and Tacrolimus as representative examples, scrutinizing the potential conditions forecasted by NMFIBC and subsequently cataloging the validation details for their top quintet of prospective illnesses. The associations of potential diseases with their respective drugs were corroborated through reputable public repositories, including DrugBank [42], CTD [43], and KEGG [44]. A synthesis of the prognostications and substantiating proof is encapsulated within Tables 4 and 5.

Levodopa, a well-established treatment for Parkinson’s disease, is highlighted in both datasets. In the original Fdataset, the relationship between Levodopa and Parkinson’s disease (168600) was not explicitly documented. However, our model was able to accurately predict this known drug–disease link, demonstrating its robust predictive power. The Fdataset also predicts additional candidate diseases such as insensitivity to pain with hyperplastic myelinopathy (147530) and restless leg syndrome (102300), with weights indicating the strength of these predictions. The Cdataset reinforces Levodopa’s association with Parkinson’s disease and introduces Alzheimer’s disease (605055) as a candidate for further investigation. Doxorubicin, primarily used in cancer treatment, shows a robust existing link with diseases like small cell cancer of the lung (182280) and breast cancer (114480) in the Fdataset. The predictions extend to other cancer types, indicating a potential broad-spectrum activity against various malignancies. The Cdataset also emphasizes Doxorubicin’s connection with cancer, predicting lymphoblastic leukemia (247640) and testicular germ cell tumor (273300) as new candidates. Amantadine, traditionally used for influenza and Parkinson’s disease, is presented with a significant existing relationship with multiple sclerosis (126200) in the Cdataset. The Fdataset predicts restless legs syndrome (102300) and Alzheimer’s disease (104300) as new candidates, suggesting a potential neuroprotective role for Amantadine beyond its current uses. Flecainide, an antiarrhythmic medication, is shown in the Cdataset with a strong existing relationship with atrial fibrillation (607554) and is predicted to have potential against dermatitis (603165) and allergic rhinitis (607154) in the Fdataset. This suggests that Flecainide may have immunomodulatory properties that could be harnessed for non-cardiac conditions. Tacrolimus, an immunosuppressive drug, is associated with asthma (208550) and dermatitis (603165) in the Fdataset, hinting at its potential in treating autoimmune and inflammatory conditions. The Cdataset also points to Tacrolimus’s potential in treating dermatitis (605805) and asthma (600807), further supporting its broad application in immunological disorders.

These case studies serve not only as illustrative examples of the model’s predictive capabilities but also as practical validations that demonstrate its biological relevance in real-world settings. While the core strength of NMFIBC lies in its systematic integration of biological similarity networks to enhance drug–disease association prediction, the case-level findings provide complementary insight into how these computational predictions translate into meaningful hypotheses for drug repurposing. Collectively, these findings underscore the potential of intelligent, context-aware modeling frameworks to

advance our understanding of pharmacological mechanisms and to support hypothesis-driven biomedical research. The integration of rigorous benchmarking with interpretable, biologically grounded predictions makes this approach especially valuable in navigating the complexity of modern therapeutic discovery.

## 5. Conclusions

These case studies serve not only as illustrative examples of the model's predictive capabilities but also as practical validations that demonstrate its biological relevance in real-world settings. While the core strength of NMFIBC lies in its systematic integration of biological similarity networks to enhance drug–disease association prediction, the case-level findings provide complementary insight into how these computational predictions can generate meaningful and testable hypotheses for drug repurposing.

Collectively, these findings underscore the potential of intelligent, context-aware modeling frameworks to advance our understanding of pharmacological mechanisms and support hypothesis-driven biomedical research. By combining rigorous performance benchmarking with biologically grounded interpretability, the NMFIBC framework contributes to bridging the gap between computational predictions and translational applications in therapeutic discovery.

**Author Contributions:** Conceptualization, Y.W. (Yangyang Wang), Y.W. (Yaping Wang) and J.W.; methodology, Y.W. (Yaping Wang) and Y.H.; validation, Y.W. (Yangyang Wang) and J.W.; investigation, Y.W. (Yangyang Wang) and Y.W. (Yaping Wang); writing—original draft preparation, Y.W. (Yangyang Wang) and Y.W. (Yaping Wang); writing—review and editing, J.W.; funding acquisition, J.W. All authors have read and agreed to the published version of the manuscript.

**Funding:** This research was funded by the Open Funds for Shaanxi Provincial Key Laboratory of Infection and Immune Diseases (No. 2023-KFMS-1); Research Project of Yan'an University (No. YAU202512552).

**Institutional Review Board Statement:** Not applicable.

**Informed Consent Statement:** Not applicable.

**Data Availability Statement:** The raw data supporting the conclusions of this article will be made available by the authors upon request.

**Conflicts of Interest:** The authors declare no conflicts of interest.

## References

1. Jourdan, J.-P.; Bureau, R.; Rochais, C.; Dallemagne, P. Drug repositioning: A brief overview. *J. Pharm. Pharmacol.* **2020**, *72*, 1145–1151. [[CrossRef](#)] [[PubMed](#)]
2. Hua, Y.; Dai, X.; Xu, Y.; Xing, G.; Liu, H.; Lu, T.; Chen, Y.; Zhang, Y. Drug repositioning: Progress and challenges in drug discovery for various diseases. *Eur. J. Med. Chem.* **2022**, *234*, 114239. [[CrossRef](#)]
3. Low, Z.Y.; Farouk, I.A.; Lal, S.K. Drug repositioning: New approaches and future prospects for life-debilitating diseases and the COVID-19 pandemic outbreak. *Viruses* **2020**, *12*, 1058. [[CrossRef](#)] [[PubMed](#)]
4. Jarada, T.N.; Rokne, J.G.; Alhaji, R. A review of computational drug repositioning: Strategies, approaches, opportunities, challenges, and directions. *J. Cheminform.* **2020**, *12*, 46. [[CrossRef](#)]
5. Yu, J.-L.; Dai, Q.-Q.; Li, G.-B. Deep learning in target prediction and drug repositioning: Recent advances and challenges. *Drug Discov. Today* **2022**, *27*, 1796–1814. [[CrossRef](#)] [[PubMed](#)]
6. Sadeghi, S.; Lu, J.; Ngom, A. A network-based drug repurposing method via non-negative matrix factorization. *Bioinformatics* **2022**, *38*, 1369–1377. [[CrossRef](#)]
7. Lee, D.D.; Seung, H.S. Learning the parts of objects by non-negative matrix factorization. *Nature* **1999**, *401*, 788–791. [[CrossRef](#)]
8. Tang, Y.; Li, G.; Wu, Y.; Yi, D. Identifying potential miRNA-disease associations based on an improved manifold learning framework. *IEEE Access* **2020**, *8*, 33263–33275. [[CrossRef](#)]

9. Wang, Y.Y.; Cui, C.; Qi, L.; Yan, H.; Zhao, X.M. DrPOCS: Drug Repositioning Based on Projection onto Convex Sets. *IEEE/ACM Trans. Comput. Biol. Bioinform.* **2019**, *16*, 154–162. [\[CrossRef\]](#)
10. Liu, J.; Zuo, Z.; Wu, G. Link prediction only with interaction data and its application on drug repositioning. *IEEE Trans. NanoBioscience* **2020**, *19*, 547–555. [\[CrossRef\]](#)
11. Luo, J.; Ding, P.; Liang, C.; Cao, B.; Chen, X. Collective prediction of disease-associated miRNAs based on transduction learning. *IEEE/ACM Trans. Comput. Biol. Bioinform.* **2016**, *14*, 1468–1475. [\[CrossRef\]](#)
12. Zhang, W.; Liu, X.; Chen, Y.; Wu, W.; Wang, W.; Li, X. Feature-derived graph regularized matrix factorization for predicting drug side effects. *Neurocomputing* **2018**, *287*, 154–162. [\[CrossRef\]](#)
13. Zhang, W.; Yue, X.; Lin, W.; Wu, W.; Liu, R.; Huang, F.; Liu, F. Predicting drug-disease associations by using similarity constrained matrix factorization. *BMC Bioinform.* **2018**, *19*, 233. [\[CrossRef\]](#)
14. Yang, M.; Luo, H.; Li, Y.; Wang, J. Drug repositioning based on bounded nuclear norm regularization. *Bioinformatics* **2019**, *35*, i455–i463. [\[CrossRef\]](#) [\[PubMed\]](#)
15. Timilsina, M.; Tandan, M.; d’Aquin, M.; Yang, H. Discovering Links Between Side Effects and Drugs Using a Diffusion Based Method. *Sci. Rep.* **2019**, *9*, 10436. [\[CrossRef\]](#)
16. Luo, H.; Wang, J.; Li, M.; Luo, J.; Peng, X.; Wu, F.X.; Pan, Y. Drug repositioning based on comprehensive similarity measures and Bi-Random walk algorithm. *Bioinformatics* **2016**, *32*, 2664–2671. [\[CrossRef\]](#) [\[PubMed\]](#)
17. Gottlieb, A.; Stein, G.Y.; Rupp, E.; Sharan, R. PREDICT: A method for inferring novel drug indications with application to personalized medicine. *Mol. Syst. Biol.* **2011**, *7*, 496. [\[CrossRef\]](#)
18. Lee, D.; Seung, H.S. Algorithms for non-negative matrix factorization. In Proceedings of the Advances in Neural Information Processing Systems 13 (NIPS 2000), Denver, CO, USA, 1 January 2000.
19. Guan, N.; Tao, D.; Luo, Z.; Yuan, B. Manifold regularized discriminative nonnegative matrix factorization with fast gradient descent. *IEEE Trans. Image Process.* **2011**, *20*, 2030–2048. [\[CrossRef\]](#)
20. Yuan, L.; Zhu, L.; Guo, W.-L.; Zhou, X.; Zhang, Y.; Huang, Z.; Huang, D.-S. Nonconvex penalty based low-rank representation and sparse regression for eQTL mapping. *IEEE/ACM Trans. Comput. Biol. Bioinform.* **2016**, *14*, 1154–1164. [\[CrossRef\]](#)
21. Cai, D.; He, X.; Han, J.; Huang, T.S. Graph regularized nonnegative matrix factorization for data representation. *IEEE Trans. Pattern Anal. Mach. Intell.* **2010**, *33*, 1548–1560.
22. Li, X.; Cui, G.; Dong, Y. Graph regularized non-negative low-rank matrix factorization for image clustering. *IEEE Trans. Cybern.* **2016**, *47*, 3840–3853. [\[CrossRef\]](#)
23. Xiao, Q.; Luo, J.; Liang, C.; Cai, J.; Ding, P. A graph regularized non-negative matrix factorization method for identifying microRNA-disease associations. *Bioinformatics* **2017**, *34*, 239–248. [\[CrossRef\]](#) [\[PubMed\]](#)
24. Facchinei, F.; Kanzow, C.; Sagratella, S. Solving quasi-variational inequalities via their KKT conditions. *Math. Program.* **2014**, *144*, 369–412. [\[CrossRef\]](#)
25. Chen, X.; Wang, L.; Qu, J.; Guan, N.N.; Li, J.Q. Predicting miRNA-disease association based on inductive matrix completion. *Bioinformatics* **2018**, *34*, 4256–4265. [\[CrossRef\]](#) [\[PubMed\]](#)
26. Gu, C.; Liao, B.; Li, X.; Li, K. Network Consistency Projection for Human miRNA-Disease Associations Inference. *Sci. Rep.* **2016**, *6*, 36054. [\[CrossRef\]](#) [\[PubMed\]](#)
27. Chen, X.; Yan, G.Y. Semi-supervised learning for potential human microRNA-disease associations inference. *Sci. Rep.* **2014**, *4*, 5501. [\[CrossRef\]](#) [\[PubMed\]](#)
28. Lu, C.; Yang, M.; Luo, F.; Wu, F.X.; Li, M.; Pan, Y.; Li, Y.; Wang, J. Prediction of lncRNA-disease associations based on inductive matrix completion. *Bioinformatics* **2018**, *34*, 3357–3364. [\[CrossRef\]](#)
29. Newson, J.M.; Santos, C.D.; Walters, B.L.; Todd, B.R. The Case of Flecainide Toxicity: What to Look for and How to Treat. *J. Emerg. Med.* **2020**, *59*, e43–e47. [\[CrossRef\]](#)
30. Araujo, G.; Nascimento, M.; Arruda, P.; Moraes, L.F.; Carvalho, M.D.A.; Medeiros, D.; Sarinho, E. The importance of tacrolimus in the treatment of allergic keratoconjunctivitis. *World Allergy Organ. J.* **2015**, *8* (Suppl. S1), A234. [\[CrossRef\]](#)
31. Taniguchi, H.; Tokui, K.; Iwata, Y.; Abo, H.; Izumi, S. A case of severe bronchial asthma controlled with tacrolimus. *J. Allergy* **2011**, *2011*, 479129. [\[CrossRef\]](#)
32. Gupta, A.; Dai, Y.; Vethanayagam, R.R.; Hebert, M.F.; Thummel, K.E.; Unadkat, J.D.; Ross, D.D.; Mao, Q. Cyclosporin A, tacrolimus and sirolimus are potent inhibitors of the human breast cancer resistance protein (ABCG2) and reverse resistance to mitoxantrone and topotecan. *Cancer Chemother. Pharmacol.* **2006**, *58*, 374–383. [\[CrossRef\]](#) [\[PubMed\]](#)
33. Nakata, S.; Kakimoto, K.; Numa, K.; Kinoshita, N.; Kawasaki, Y.; Tatsumi, Y.; Tawa, H.; Koshihara, R.; Hirata, Y.; Ota, K.; et al. Risk Factors for Nephrotoxicity due to Tacrolimus Therapy for Ulcerative Colitis. *Digestion* **2022**, *103*, 339–346. [\[CrossRef\]](#) [\[PubMed\]](#)
34. Molloy, S.; McKeith, I.G.; O’Brien, J.T.; Burn, D.J. The role of levodopa in the management of dementia with Lewy bodies. *J. Neurol. Neurosurg. Psychiatry* **2005**, *76*, 1200–1203. [\[CrossRef\]](#)
35. Vaculova, A.; Kaminsky, V.; Jalalvand, E.; Surova, O.; Zhivotovsky, B. Doxorubicin and etoposide sensitize small cell lung carcinoma cells expressing caspase-8 to TRAIL. *Mol. Cancer* **2010**, *9*, 87. [\[CrossRef\]](#) [\[PubMed\]](#)

36. Manchun, S.; Dass, C.R.; Cheewatanakornkool, K.; Sriamornsak, P. Enhanced anti-tumor effect of pH-responsive dextrin nanogels delivering doxorubicin on colorectal cancer. *Carbohydr. Polym.* **2015**, *126*, 222–230. [[CrossRef](#)]
37. Habas, K.; Anderson, D.; Brinkworth, M.H. Germ cell responses to doxorubicin exposure in vitro. *Toxicol. Lett.* **2017**, *265*, 70–76. [[CrossRef](#)] [[PubMed](#)]
38. Dai, S.; Ye, Z.; Wang, F.; Yan, F.; Wang, L.; Fang, J.; Wang, Z.; Fu, Z. Doxorubicin-loaded poly( $\epsilon$ -caprolactone)-Pluronic micelle for targeted therapy of esophageal cancer. *J. Cell. Biochem.* **2018**, *119*, 9017–9027. [[CrossRef](#)]
39. Evidente, V.G.H.; Adler, C.H.; Caviness, J.N.; Hentz, J.G.; Gwinn-Hardy, K. Amantadine is beneficial in restless legs syndrome. *Mov. Disord.* **2000**, *15*, 324–327. [[CrossRef](#)]
40. Andrikopoulos, G.K.; Pastromas, S.; Tzeis, S. Flecainide: Current status and perspectives in arrhythmia management. *World J. Cardiol.* **2015**, *7*, 76–85. [[CrossRef](#)]
41. Canzanello, V.J.; Textor, S.C.; Taler, S.J.; Schwartz, L.L.; Porayko, M.K.; Wiesner, R.H.; Krom, R.A. Late hypertension after liver transplantation: A comparison of cyclosporine and tacrolimus (FK 506). *Liver Transplant. Surg.* **1998**, *4*, 328–334. [[CrossRef](#)]
42. Wishart, D.S.; Feunang, Y.D.; Guo, A.C.; Lo, E.J.; Marcu, A.; Grant, J.R.; Sajed, T.; Johnson, D.; Li, C.; Sayeeda, Z.; et al. DrugBank 5.0: A major update to the DrugBank database for 2018. *Nucleic Acids Res.* **2018**, *46*, 1074–1082. [[CrossRef](#)] [[PubMed](#)]
43. Davis, A.P.; Wiegiers, T.C.; Johnson, R.J.; Sciaky, D.; Wiegiers, J.; Mattingly, C.J. Comparative Toxicogenomics Database (CTD): Update 2023. *Nucleic Acids Res.* **2023**, *51*, D1257–D1262. [[CrossRef](#)] [[PubMed](#)]
44. Kanehisa, M.; Goto, S.; Sato, Y.; Kawashima, M.; Furumichi, M.; Tanabe, M. Data, information, knowledge and principle: Back to metabolism in KEGG. *Nucleic Acids Res.* **2014**, *42*, D199–D205. [[CrossRef](#)]

**Disclaimer/Publisher’s Note:** The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.