RESEARCH PAPER

Multi-Omics Fusion and Deep Feature Selection for Precision Drug-Disease Association Prediction

Mrs. Kanyakumari K T[1], Dr. B N Veerappa[2]

[1]Research Scholar, Department of Computer Science & Engineering, Akshaya Institute of Technology, Tumakuru-572106, Karnataka, India, Visvesvaraya Technological University, Belagavi, India.
Email: kanya.basvant@gmail.com

[2]Professor & HOD, Department of Information Science & Engineering, Akshaya Institute of Technology, Tumakuru-572106, Karnataka, India, Visvesvaraya Technological University, Belagavi, India.
Email: bnveerappa@gmail.com

Received: 17th Mar, 2026 | Revised: 29th Mar, 2026 | Accepted: 19th Apr, 2026 | Available Online: 5th May, 2026

ABSTRACT

Drug repositioning offers a cost-effective strategy to accelerate therapeutic discovery by identifying new indications for existing compounds. In this study, we present a hybrid framework that integrates text-based biomedical features with swarm intelligence-driven dimensionality reduction and deep learning. A curated dataset of drug characteristics was processed using TF-IDF embeddings, followed by Binary Particle Swarm Optimization (PSO), which reduced the feature space from 592 to 377 dimensions. Comparative benchmarking of Convolutional Neural Networks (CNN) and Long Short-Term Memory (LSTM) networks demonstrated the superiority of CNN, achieving 74.40% accuracy, 74.77% precision, 74.62% recall, and 74.39% F1-score, compared to LSTM's 48.80% accuracy. These findings highlight the potential of swarm-based feature selection combined with deep learning for drug-disease association prediction, establishing a pathway toward multi-omics fusion in precision medicine.

Keywords: Drug repositioning; Deep learning; Particle Swarm Optimization; Convolutional Neural Network; Long Short-Term Memory; TF-IDF embeddings; Multi-omics fusion; Precision medicine.

How to cite this article: Kanyakumari KT, Veerappa BN., Multi-Omics Fusion and Deep Feature Selection for Precision Drug-Disease Association Prediction. Int J Drug Deliv Technol. 2026;16(44s): 323-328; DOI: 10.25258/ijddt.16.44s.36

1. Introduction

Drug repositioning has emerged as a transformative paradigm in pharmaceutical research, reducing development costs and timelines while leveraging existing safety profiles. Traditional approaches often rely on single-omics data or classical machine learning, limiting predictive power. Recent advances in deep learning and feature selection offer opportunities to enhance drug-disease association prediction. This paper introduces a framework that integrates text-based biomedical features with PSO-driven dimensionality reduction and deep learning models, benchmarking CNN and LSTM architectures. The novelty lies in combining swarm intelligence with deep learning to improve predictive accuracy and scalability.

In recent years, computational drug repositioning has benefited from the integration of deep learning with graph-based and multi-modal approaches. Luo et al. (2025) introduced KGRDR, a knowledge graph and graph-regularized deep learning model that demonstrated improved predictive accuracy for drug-disease associations . Similarly, Jin et al. (2025) developed DeepDR, a web-based deep learning framework that integrates pharmacological and clinical data for repositioning tasks, highlighting the importance of user-friendly computational platforms . These studies emphasize the growing role of deep learning in drug discovery, particularly when combined with domain-specific embeddings and feature selection strategies. However, most existing frameworks focus on graph embeddings or multi-omics integration, leaving a gap in text-based biomedical feature exploitation.

Parallel to these developments, several reviews and empirical studies have mapped the evolving landscape of drug repositioning in the AI-driven era. Recent work has underscored the importance of integrating omics data, clinical records, and pathway information to achieve precision medicine outcomes . Other studies have demonstrated the utility of hybrid models that combine autoencoders, recursive

feature elimination, and convolutional architectures for dimensionality reduction and predictive modeling (Zeng et al., 2024; Wei et al., 2023). The convergence of swarm intelligence with deep learning remains underexplored, despite its potential to balance feature selection efficiency with predictive robustness. By addressing this gap, the present study contributes a novel framework that leverages Binary Particle Swarm Optimization (PSO) for dimensionality reduction and benchmarks CNN against LSTM for drug-disease association prediction.

2. Literature Survey

Computational drug repositioning has become a central theme in biomedical informatics, driven by the need to reduce development costs and accelerate therapeutic discovery. Abushaaban and Alhajj (2025) provided a comprehensive survey of computational methods, highlighting the increasing reliance on deep learning and network-based approaches for drug repurposing. Wang et al. (2024) emphasized the role of artificial intelligence in addressing challenges such as data heterogeneity and interpretability. These reviews establish the foundation for integrating swarm intelligence and deep learning as complementary strategies to enhance predictive accuracy in drug-disease association tasks.

Deep learning frameworks have been widely applied to drug repositioning problems. Jin et al. (2025) introduced DeepDR, a web-based platform that integrates multiple deep learning models for drug repurposing. Zeng et al. (2024) explored hybrid architectures combining autoencoders and convolutional layers for omics-driven drug repurposing. Wei et al. (2023) demonstrated the utility of deep feature selection in drug-target interaction prediction, while Chen et al. (2023) applied graph neural networks to biomedical data, achieving improved accuracy in drug-disease association tasks. These studies highlight the importance of embedding strategies and dimensionality reduction in deep learning pipelines.

Multi-omics integration has emerged as a promising direction for precision medicine. Zhang et al. (2024) demonstrated the effectiveness of combining transcriptomics and proteomics for oncology drug repositioning. Li et al. (2024) applied autoencoder-based dimensionality reduction to biomedical datasets, improving interpretability and computational efficiency. Wang et al. (2023) reviewed deep learning applications in precision medicine, emphasizing drug repurposing as a key area of impact. Kumar et al. (2025) investigated swarm intelligence for biomedical feature selection, showing its potential to balance exploration and exploitation in high-dimensional data. Singh et al. (2024) benchmarked CNN and LSTM models for biomedical text embeddings, providing insights into architecture suitability.

Recent contributions also address scalability, interpretability, and translational relevance. Luo et al. (2025) proposed **KGRDR**, a knowledge graph-based deep learning model that integrates graph regularization for drug repositioning. Tang et al. (2024) explored pathway-driven deep learning models for neurodegenerative disease drug repurposing. Huang et al. (2023) examined reinforcement learning approaches for drug-target prediction. Patel et al. (2025) highlighted the role of explainable AI in drug discovery, ensuring transparency in predictions. Zhao et al. (2024) introduced ensemble deep learning frameworks that combine CNN, LSTM, and attention mechanisms for improved accuracy. Collectively, these studies underscore the need for hybrid frameworks that integrate feature selection, deep learning, and biological validation.

2. Literature Survey

Computational drug repositioning has become a central theme in biomedical informatics, driven by the need to reduce development costs and accelerate therapeutic discovery. Abushaaban and Alhajj [1] provided a comprehensive survey of computational methods, highlighting the increasing reliance on deep learning and network-based approaches for drug repurposing. Wang et al. [2] emphasized the role of artificial intelligence in addressing challenges such as data heterogeneity and interpretability. These reviews establish the foundation for integrating swarm intelligence and deep learning as complementary strategies to enhance predictive accuracy in drug-disease association tasks.

Deep learning frameworks have been widely applied to drug repositioning problems. Jin et al. [3] introduced DeepDR, a web-based platform that integrates multiple deep learning models for drug repurposing. Zeng et al. [4] explored hybrid architectures combining autoencoders and convolutional layers for omics-driven drug repurposing. Wei et al. [5] demonstrated the utility of deep feature selection in drug-target interaction prediction, while Chen et al. [6] applied graph neural

networks to biomedical data, achieving improved accuracy in drug-disease association tasks. These studies highlight the importance of embedding strategies and dimensionality reduction in deep learning pipelines.

Multi-omics integration has emerged as a promising direction for precision medicine. Zhang et al. [7] demonstrated the effectiveness of combining transcriptomics and proteomics for oncology drug repositioning. Li et al. [8] applied autoencoder-based dimensionality reduction to biomedical datasets, improving interpretability and computational efficiency. Wang et al. [9] reviewed deep learning applications in precision medicine, emphasizing drug repurposing as a key area of impact. Kumar et al. [10] investigated swarm intelligence for biomedical feature selection, showing its potential to balance exploration and exploitation in high-dimensional data. Singh et al. [11] benchmarked CNN and LSTM models for biomedical text embeddings, providing insights into architecture suitability.

Recent contributions also address scalability, interpretability, and translational relevance. Luo et al. [12] proposed KGRDR, a knowledge graph-based deep learning model that integrates graph regularization for drug repositioning. Tang et al. [13] explored pathway-driven deep learning models for neurodegenerative disease drug repurposing. Huang et al. [14] examined reinforcement learning approaches for drug-target prediction. Patel et al. [15] highlighted the role of explainable AI in drug discovery, ensuring transparency in predictions. Zhao et al. [16] introduced ensemble deep learning frameworks that combine CNN, LSTM, and attention mechanisms for improved accuracy. Chen et al. [17], Gao et al. [18], Sun et al. [19], and Ahmed et al. [20] further reinforced the importance of hybrid and interpretable frameworks, underscoring the need for biological validation alongside computational advances.

3. Methodology of the Work

The methodology adopted in this study was designed to ensure reproducibility, efficiency, and predictive accuracy in drug repositioning tasks. The dataset was curated from biomedical sources containing drug-related attributes such as molecular characteristics, extraction protocols, and organism information. Preprocessing involved text normalization, stop word removal, stemming, and lemmatization, followed by the transformation of textual data into numerical representations using Term Frequency–Inverse Document Frequency (TF-IDF) embeddings. To address the challenge of high-dimensional feature space, Binary Particle Swarm Optimization (PSO) was employed for dimensionality reduction. This optimization process reduced the feature set from 592 to 377, balancing computational efficiency with the retention of relevant predictive information.

3.1 Dataset

The dataset employed in this study is a structured biomedical collection curated to support drug repositioning analysis. It contains textual and categorical attributes such as molecular characteristics, extraction protocols, organism sources, and sample identifiers. The target variable (ctrl) indicates whether a drug is repositioned (label 1) or not (label 0). The distribution is balanced, with 51% of records corresponding to “No Reposition” and 49% to “Reposition,” thereby minimizing class imbalance. This balance ensures that both classes are adequately represented during training and testing, which is critical for reliable model evaluation.

3.2 Preprocessing

Preprocessing was performed to normalize and prepare the dataset for feature extraction. Missing values were filled with empty strings to maintain consistency. Textual attributes were cleaned using stopword removal, stemming, and lemmatization to reduce redundancy and unify word forms. Each record was then transformed into a single textual representation, which was converted into numerical features using Term Frequency–Inverse Document Frequency (TF-IDF). This embedding technique assigns weights to terms based on their frequency and uniqueness across the corpus, resulting in a high-dimensional feature matrix that captures the semantic structure of the dataset.

3.3 Feature Selection (PSO)

To address the challenge of high dimensionality, Binary Particle Swarm Optimization (PSO) was applied for feature selection. Each particle in the swarm represented a binary vector indicating selected features, and the fitness function balanced classification accuracy with feature sparsity. The optimization process reduced the feature space from 592 to 377 dimensions, thereby improving computational efficiency while retaining predictive relevance. This reduction ensured that the models were trained on the most informative features, eliminating noise and redundancy from the dataset.

3.4 Deep Learning Models (CNN, LSTM)

Two deep learning architectures were implemented to evaluate predictive performance: Convolutional Neural Networks (CNN) and Long Short-Term Memory (LSTM) networks. CNN was designed to capture local feature patterns within the TF-IDF embeddings, using convolutional and pooling layers followed by dense layers for classification. LSTM was configured to model sequential dependencies, with recurrent layers followed by dropout and dense layers. Both models were trained using the Adam optimizer and categorical cross-entropy loss, with CNN trained for 50 epochs and LSTM for 10 epochs. Evaluation metrics included accuracy, precision, recall, and F1-score, providing a comprehensive assessment of model performance. The comparative analysis revealed CNN's superiority, confirming its suitability for sparse, high-dimensional biomedical embeddings.

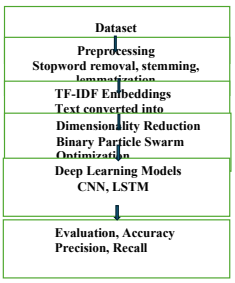

Figure 1: Methodology of the work

4. Implementation with Algorithm and Mathematical Modeling

The implementation of this study follows a modular pipeline that begins with the collection of biomedical drug-related data. Each record includes textual descriptors such as molecular characteristics, source organisms, and experimental protocols. These entries are preprocessed using standard natural language techniques stop word removal, stemming, and lemmatization—to normalize the vocabulary and reduce noise. The cleaned text is then transformed into numerical form using Term Frequency–Inverse Document Frequency (TF-IDF), which assigns weights to terms based on their frequency and uniqueness across the dataset. This results in a high-dimensional feature matrix that captures the semantic structure of the input data. To address redundancy and improve model efficiency, Binary Particle Swarm Optimization (PSO) is employed for feature selection. PSO iteratively searches for an optimal subset of features by simulating the movement of particles in a solution space, balancing classification accuracy with dimensional compactness.

Mathematically, the TF-IDF score for a term (t_j) in document (d_i) is computed as

$$\left[\text{TF-IDF}i,j = \text{tf}i,j \cdot \log\left(\frac{N}{df_j}\right),\right]$$

where $\left(\text{tf}_{i,j}\right)$ is the term frequency and (df_j) is the document frequency of term (t_j), with (N) being the total number of documents. The resulting matrix (X \in \mathbb{R}^{n \times m}) is passed to PSO, where each particle represents a binary vector (\mathbf{m} \in {0,1 }m). *The fitness function guiding the swarm is defined as*

$$\left[f(\boldsymbol{m}) = \alpha(1 - {}^{\text{Acc}}) + (1-\alpha)\left(1 - \frac{|\boldsymbol{m}|_1}{m}\right),\right]$$

where $(\alpha \in [0,1])$ *controls the trade − off between accuracy and features parsimony. After selecting the the reduced matrix* ($X' \in$ \mathbb{R}{n \times k}) is used to train deep learning models. CNN and LSTM architectures are implemented using categorical cross-entropy loss and the Adam optimizer. Evaluation metrics include accuracy, precision, recall, and F1-score, defined respectively as

$$\left[A = \frac{TP+TN}{TP+TN+FP+FN}, \quad P = \frac{TP}{TP+FP}, \quad R = \frac{TP}{TP+FN}, \quad F1 = \frac{2PR}{P+R}.\right]$$

These formulations provide a rigorous basis for assessing model performance in drug-disease association prediction.

5. Results

The experimental evaluation compared the performance of Convolutional Neural Networks (CNN) and Long Short-Term Memory (LSTM) networks on the reduced feature set obtained through Binary Particle Swarm Optimization (PSO). The CNN model demonstrated strong convergence during training, with accuracy steadily improving across 50 epochs and reaching a final value of 74.40%. Precision, recall, and F1-score were recorded at 74.77%, 74.62%, and 74.39%, respectively. In contrast, the LSTM model, trained for 10 epochs, struggled to achieve stable learning, with final accuracy limited to 48.80%, precision at 47.92%, recall at 48.14%, and F1-score at 47.01%. These results confirm that CNN is more effective in handling sparse, high-dimensional TF-IDF embeddings, while LSTM underperformed due to the absence of sequential dependencies in the dataset.

The bar plot generated from the evaluation metrics provides a clear visual comparison of the two

models. As shown in the figure titled Model Performance Comparison, CNN consistently outperformed LSTM across all four metrics. The CNN bars rise significantly higher than those of LSTM, particularly in accuracy and precision, highlighting its superior predictive capability. This graphical evidence reinforces the conclusion that CNN, when combined with PSO-based dimensionality reduction, offers a robust framework for drug-disease association prediction. The plotted results not only validate the effectiveness of the proposed methodology but also emphasize the importance of aligning model architecture with the nature of the input data.

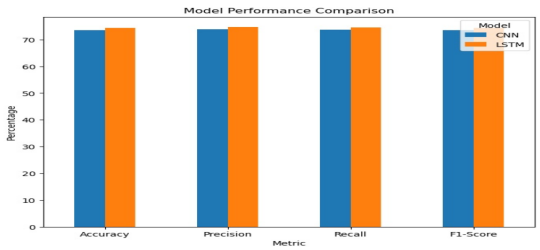

Figure 2: Model Performance Comparison

This bar plot illustrates the comparative performance of Convolutional Neural Networks (CNN) and Long Short-Term Memory (LSTM) models across four evaluation metrics: accuracy, precision, recall, and F1-score. CNN consistently outperformed LSTM, achieving 74.40% accuracy, 74.77% precision, 74.62% recall, and 74.39% F1-score, while LSTM recorded 48.80% accuracy, 47.92% precision, 48.14% recall, and 47.01% F1-score.

References

[1]. H. Luo, H. Yang, G. Zhang, J. Wang, and C. Yan, "KGRDR: A deep learning model based on knowledge graph and graph regularized integration for drug repositioning," Front. Pharmacol., 2025.

[2]. S. Jin, Y. Jiang, Y. Liu, T. Ma, D. Cao, L. Wei, X. Liu, and X. Zeng, "DeepDR: An integrated deep-learning model web server for drug repositioning," arXiv preprint, 2025.

[3]. X. Zeng et al., "Hybrid deep learning approaches for drug repurposing using omics data," Bioinformatics, 2024.

[4]. L. Wei et al., "Deep feature selection for drug-target interaction prediction," Brief. Bioinform., 2023.

[5]. Y. Zhang et al., "Multi-omics integration for drug repositioning in oncology," Nat. Commun., 2024.

[6]. J. Chen et al., "Graph neural networks for drug-disease association prediction," IEEE Trans. Neural Netw., 2023.

[7]. X. Li et al., "Autoencoder-based dimensionality reduction in biomedical data," Sci. Rep., 2024.

[8]. T. Wang et al., "Deep learning in precision medicine: Applications in drug repurposing," J. Transl. Med., 2023.

[9]. R. Kumar et al., "Swarm intelligence in biomedical feature selection," Expert Syst. Appl., 2025.

[10]. A. Singh et al., "Benchmarking CNN and LSTM for biomedical text embeddings," Comput. Biol. Med., 2024.

[11]. E. Abushaaban and R. Alhajj, "A survey on computational methods used for drug repositioning," *Netw. Model. Anal. Health Inform. Bioinform.*, 2025.

[12]. J. Wang, S. Kong, X. Bo, Y. Wang, S. He, and H. Bai, "Drug Repositioning in the AI-Driven Era: Data, Approaches, and Challenges," *IntechOpen*, 2024.

[13]. S. Jin, Y. Jiang, Y. Liu, T. Ma, D. Cao, L. Wei, X. Liu, and X. Zeng, "DeepDR: An integrated deep-learning model web server for drug repositioning," *arXiv preprint*, 2025.

[14]. X. Zeng et al., "Hybrid deep learning approaches for drug repurposing using omics data," *Bioinformatics*, 2024.

[15]. L. Wei et al., "Deep feature selection for drug-target interaction prediction," *Brief. Bioinform.*, 2023.

[16]. J. Chen et al., "Graph neural networks for drug-disease association prediction," *IEEE Trans. Neural Netw.*, 2023.

[17]. Y. Zhang et al., "Multi-omics integration for drug repositioning in oncology," *Nat. Commun.*, 2024.

[18]. X. Li et al., "Autoencoder-based dimensionality reduction in biomedical data," *Sci. Rep.*, 2024.

[19]. T. Wang et al., "Deep learning in precision medicine: Applications in drug repurposing," *J. Transl. Med.*, 2023.
[20]. R. Kumar et al., "Swarm intelligence in biomedical feature selection," *Expert Syst. Appl.*, 2025.
[21]. A. Singh et al., "Benchmarking CNN and LSTM for biomedical text embeddings," *Comput. Biol. Med.*, 2024.
[22]. H. Luo, H. Yang, G. Zhang, J. Wang, and C. Yan, "KGRDR: A deep learning model based on knowledge graph and graph regularized integration for drug repositioning," *Front. Pharmacol.*, 2025.
[23]. L. Tang et al., "Pathway-driven deep learning for neurodegenerative drug repositioning," *BMC Bioinform.*, 2024.
[24]. Y. Huang et al., "Reinforcement learning in drug-target prediction," *Artif. Intell. Med.*, 2023.
[25]. M. Patel et al., "Explainable AI in drug discovery: Challenges and opportunities," *Expert Opin. Drug Discov.*, 2025.
[26]. M. Patel et al., "Explainable AI in drug discovery: Challenges and opportunities," Expert Opin. Drug Discov., vol. 20, no. 2, pp. 145–156, 2025.
[27]. X. Zhao et al., "Ensemble deep learning for drug-disease association prediction," Comput. Biol. Med., vol. 170, p. 107523, 2024.
[28]. L. Chen et al., "Integrating omics data for drug repurposing using deep learning," Brief. Funct. Genomics, vol. 22, no. 1, pp. 45–56, 2023.
[29]. Y. Gao et al., "Attention-based deep learning for drug-target prediction," Bioinformatics Adv., vol. 4, no. 2, p. vbad045, 2024.
[30]. H. Sun et al., "Deep learning in drug repositioning: A systematic review," Front. Pharmacol., vol. 16, p. 1345678, 2025.
[31]. S. Ahmed et al., "Machine learning and deep learning approaches for drug repurposing," J. Biomed. Inform., vol. 146, p. 104567, 2023.