INTERNATIONAL JOURNAL OF ADVANCE RESEARCH, IDEAS AND INNOVATIONS IN TECHNOLOGY
IJARIIT
ISSN: 2454-132X
Impact Factor: 6.078
(Volume 12, Issue 1 - V12I1-1157)
Available online at: https://www.ijariit.com

A Survey on Modern Computational Methods for Drug Repurposing

Aditi Dipak Thorat
aditi.thorat2004@gmail.com
Pune Institute of Computer Technology, Maharashtra

Shlok Shivaji Kaule
shlokkaule1707@gmail.com
Pune Institute of Computer Technology, Maharashtra

Paras Vijay Tak
parastak429@gmail.com
Pune Institute of Computer Technology, Maharashtra

Anuj Prakash Gagare
anujgagare18@gmail.com
Pune Institute of Computer Technology, Maharashtra

Vijayendra S. Gaikwad
vsgaikwad@ms.pict.edu
Pune Institute of Computer Technology, Maharashtra

ABSTRACT

Drug repurposing, the process of identifying new therapeutic uses for existing drugs, offers a promising strategy to accelerate drug development by significantly reducing costs, time, and risks compared to de novo drug discovery. The increasing availability of large-scale biomedical data has catalysed the development of computational approaches to systematically identify and prioritise repurposing candidates. This survey reviews the state-of-the-art computational methodologies, with a particular focus on network medicine and machine learning-based techniques. We discuss key approaches such as pathway-based analysis, network proximity, matrix factorisation, and the growing application of deep learning, particularly Graph Neural Networks (GNNs), which leverage complex biomedical networks. The paper explores how these methodsutilisee heterogeneous data—including drug-target interactions, gene-disease associations, and molecular structures—to generate repurposing hypotheses. Furthermore, we outline the primary challenges in the field, including data integration, model generalizability, and the need for explainability, and discuss future directions, such as the integration of multi-modal data and the development of more sophisticated, interpretable AI models.

Keywords: *Computational Drug Discovery, Graph Attention Networks, Network-Based Prediction, Heterogeneous Graphs, Machine Learning, Therapeutic Discovery.*

I. INTRODUCTION

The development of new pharmaceuticals, or de novo drug discovery, is a notoriously challenging process. It is characterized by immense costs, protracted timelines, and a high risk of failure. On average, bringing a single new drug to market can take 10-15 years and cost upwards of $2.5 billion [1]. A significant portion of this cost and time is consumed by extensive preclinical testing and multi-phase clinical trials required to establish the safety and efficacy of a new molecular entity. The attrition rate is daunting; the vast majority of candidate molecules fail during development, with many failing in latestage clinical trials after hundreds of millions of dollars have already been invested [2]. This inefficient paradigm creates significant barriers to addressing unmet medical needs, particularly for rare diseases where small patient populations make it difficult to recoup development costs, and for emerging public health crises that demand rapid therapeutic solutions [1], [3].

In response to these challenges, drug repurposing (also known as drug repositioning) has emerged as a critical and effective alternative strategy. Drug repurposing is the process of identifying new therapeutic uses for drugs that have already been approved for other indications [1]. Because these drugs have already undergone extensive safety and pharmacokinetic profiling, they have a well-established safety record in humans. This significantly de-risks the development process, allowing repurposed drugs to bypass early-stage clinical trials and enter directly into Phase II or III trials for the new indication. Consequently, the timeline for development can be reduced to as little as $3 - 12$ years, and the associated costs can be a fraction of those for a de novo drug [9].

Historically, many successful instances of drug repurposing were the result of serendipity-fortuitous clinical observations of unexpected side effects. Prominent examples include thalidomide, which was repurposed for treating multiple myeloma after being withdrawn for its teratogenic effects. Another well-known case is minoxidil, an antihypertensive medication that was found to promote hair growth, leading to its second life as a topical treatment for baldness. Similarly, aspirin, originally an analgesic, is now widely used in low doses as an antiplatelet agent to prevent cardiovascular events. However, relying on chance is not a sustainable strategy for modern medicine. The contemporary need is for systematic, predictable, and scalable methods to identify repurposing opportunities, especially for complex, multifactorial diseases that do not follow the traditional "one-drug-one-gene-one-disease" model [4].

The shift from serendipitous to systematic drug repurposing has been driven by a revolution in data and computation. The explosion of "big data" in biomedicine-including largescale chemical genomics, proteomics, transcriptomics, and vast databases of gene-disease associations-has provided the raw material for building comprehensive models of human biology [5], [6]. This data-rich environment has paved the way for the application of sophisticated computational techniques rooted in network medicine and artificial intelligence. These methods model the complex web of interactions between drugs, protein targets, genes, and diseases to predict therapeutic effects that are not immediately obvious [7], [8], [4]. By analyzing these intricate networks, computational approaches can generate high-quality, testable hypotheses for drug repurposing at a scale and speed unattainable through traditional experimental methods alone.

This survey provides an overview of these prominent computational methodologies, with a focus on network-based and deep learning approaches, and discusses the current challenges and future trends that will continue to shape this dynamic field.

II. RELATED WORKS

The computational drug repurposing landscape is diverse, with methods ranging from network analysis to advanced deep learning models. These approaches leverage the principle that relationships between biological entities can be computationally inferred to predict novel therapeutic uses for drugs.

A. Network Medicine and Pathway-Based Approaches

Network medicine provides a powerful framework for drug repurposing by modeling the complex interplay between diseases, genes, and drugs within a biological network, typically a protein-protein interaction (PPI) network. The central idea is that diseases are not caused by single-molecule defects but by perturbations in a complex network of interactions.

Proximity and Pathway Analysis: A key concept is the disease module, a localized network neighborhood of proteins associated with a specific disease. The therapeutic effect of a drug is often correlated with the proximity of its protein targets (T) to the corresponding disease module (S). This proximity is frequently quantified by the "closest distance" $d_c(S,T)$, which measures the average shortest path length from each drug target to the nearest disease protein in the network [9], [7]. It is formally defined as:

$$d_c(S,T) = \frac{1}{|T|}\sum_{t\in T} \min_{s\in S} d(s,t) \tag{1}$$

where $d(s,t)$ is the shortest path length between protein s and target t. To assess statistical significance, this distance is often converted into a Z-score by comparing it to a reference distribution of distances between random sets of proteins of the same size and degree:

$$Z = \frac{d_c(S,T)_{\text{observed}} - \mu_{d_c(\text{random})}}{\sigma_{d_c(\text{random})}} \tag{2}$$

A significantly negative Z-score indicates that the drug's targets are closer to the disease module than expected by chance, suggesting a potential therapeutic relationship. Marín Tercero et al. [7] applied this principle to schizophrenia, while Otero-Carrasco et al. [9] used it to prioritize drugs based on their proximity to disease-associated biological pathways.

Pattern Analysis in Specific Contexts: Computational methods are also used to analyze existing repurposing successes to uncover underlying biological patterns. In the context of rare diseases, Otero-Carrasco et al. [9] studied orphan drugs and found that repurposing often occurs between two different rare diseases. Their analysis revealed that diseases successfully treated by the same repurposed drug tend to exhibit high phenotypic similarity (i.e., share a significant number of symptoms), suggesting that shared clinical manifestations can be a strong indicator for repurposing potential.

Explainable Network Models: To address the "black box" nature of some computational methods, certain models prioritize explainability. Castiglione et al. [11] developed an approach based on biased random walks over a knowledge graph of drug-gene-disease associations. By modeling the system as an ergodic Markov process, the method not only recommends drugs but can also reveal the most probable paths through the network that led to the recommendation. The recommendation score for a drug-disease pair after an l-step walk is derived from the l^{th} power of the transition matrix, making the process transparent and traceable.

Multi-layered Recommender Systems: More complex systems integrate multiple data types into a single framework. Wang et al. [7] developed ANTENNA, a multi-layered recommender system that models drugs, genes, and diseases in distinct but interconnected network layers. It uses a tri-factorization-based collaborative filtering algorithm to infer novel genome-wide chemical-gene associations. This information is then integrated with a Random Walk with Restart (RWR) algorithm to predict and assess the statistical reliability of novel drug-disease associations, leading to the successful identification of diazoxide as a potential therapy for triplenegative breast cancer.

B. Machine Learning and Deep Learning Methods

With the growth of large-scale biomedical data, machine learning and deep learning have become central to drug repurposing.

Matrix Factorization: These techniques, common in recommender systems, treat the problem as completing a sparse matrix of known drug-disease associations. Ceddia et al. [2] proposed an innovative method based on Non-negative Matrix Tri-Factorization (NMTF). Given an association matrix R_{AB} between drugs (A) and diseases (B), NMTF approximates it as the product of three lower-dimensional matrices:

$$R_{AB} \approx \hat{R}_{AB} = G_A S_{AB} G_B^T \tag{3}$$

The reconstructed matrix $\hat{R}_{AB}$ contains predicted scores for unknown associations. A key contribution of their work is an enhancement strategy where they enrich the input matrix by using a shortest-path analysis on the PPI network to infer novel, indirect drug-protein interactions. The weight of these inferred associations is decayed exponentially with path length $|P_{ab}|$, using a factor α :

$$R'_{AB}[a,b] = \alpha^{|P_{ab}|-1} \tag{4}$$

This allows the model to leverage broader network context and significantly improves prediction performance.

Graph Neural Networks (GNNs): GNNs are purpose-built for learning from graph-structured data and have become a leading methodology for drug repurposing. They operate via a "message passing" scheme, where each node (e.g., a drug or protein) iteratively updates its feature vector (embedding) by aggregating information from its neighbors. A simplified representation of a single GNN layer is:

$$h_v^{(l+1)} = \text{UPDATE}^{(l)}\left(h_v^{(l)}, \text{AGGREGATE}^{(l)}\left(\left\{h_u^{(l)} : u \in N(v)\right\}\right)\right) \quad (5)$$

where $h_v^{(l)}$ is the embedding of node v at layer l, and the UPDATE/AGGREGATE functions are learnable neural networks. After several layers, the final embeddings capture both the node's initial features and its topological environment. Artiñano-Muñoz et al. [10] (DRAGON) and Ayuso-Muñoz et al. [6] demonstrate the power of initializing this process with rich features $h_v^{(0)}$, using embeddings from drug molecular structures and protein sequences to improve the accuracy of predicting new drug-disease links.

Advanced GNN Architectures: Researchers are developing more sophisticated GNNs to better reflect biological complexity. Moving beyond the single-drug/single-protein paradigm, Bacciu et al. [3] designed a deep graph network that predicts a drug's efficacy against a set of multiple protein targets considered jointly. This approach better captures the reality that drugs often act on functionally related protein ensembles. Li and Hu [11] proposed a hybrid framework that combines a multimodal GNN (operating on a heterogeneous graph of drugs, proteins, diseases, and pathways) with a parallel structural learning module that processes molecularlevel information directly. The predictions from both modules are then ensembled for a more robust result. Park et al. [12], while focused on predicting side-effect frequency, introduce a "dual representation learning" technique that embeds both drugs and diseases into a common vector space, where the prediction score is a function of the similarity (e.g., dot product) between their embeddings:

$$\text{score} \propto \text{embedding}_{drug} \cdot \text{embedding}_{disease} \quad (6)$$

C. Feature Representation and Data Sources

The performance of these computational models is highly dependent on the quality of input data and feature representations. Drug features are often derived from their chemical structures using tools like RDKit [13] to generate extendedconnectivity fingerprints (ECFPs) [14]. For proteins, features can be derived from amino acid sequences, with advanced models using embeddings from large language models trained on biological sequences, such as ESM [15]. These features are integrated with large-scale knowledge bases like DisGeNET [6], which consolidates information on gene-disease associations, and comprehensive platforms like DRIVE [16], which provide a unified interface for disease visualization and running various repurposing algorithms.

D. Dominant Methodologies in Computational Drug Repurposing

The reviewed literature reveals a clear trend towards integrated, data-driven methodologies that model complex biological systems. Two dominant themes emerge: the centrality of network-based representations and the increasing sophistication of machine learning models to learn from them.

Network-Based Inference: The foundational methodology across many modern approaches is the representation of biomedical knowledge as a graph. This includes PPI networks, drug-target networks, and heterogeneous graphs linking multiple entity types (drugs, genes, diseases, pathways) [9], [11], [3]. Key analytical techniques performed on these networks include:

i. Proximity and Distance Metrics: Calculating the shortest path length between drug targets and diseaseassociated genes in a PPI network is a common and effective measure of therapeutic potential [7], [7].
ii. Random Walks: Algorithms based on random walks explore network topology to rank potential drug-disease associations, with some methods providing explainability by tracing the most probable paths [11].
iii. Link Prediction: The task is often framed as predicting missing links (edges) between drug and disease nodes in a heterogeneous graph.
iv. Deep Learning on Graphs: GNNs have become the state-of-the-art for link prediction tasks in this domain [6], [11]. Their primary advantage is the ability to learn node embeddings that capture both the node's intrinsic properties (e.g., a drug's chemical structure) and its topological context within the broader network. The most successful recent models tend to be multimodal, integrating diverse feature types such as molecular graphs, protein sequences, and pathway information into the GNN architecture [11], [12].

III. PROPOSED METHODOLOGY

The reviewed literature reveals a clear trend towards integrated, data-driven methodologies that model complex biological systems. Two dominant themes emerge: the centrality of network-based representations and the increasing sophistication of machine learning models to learn from them. Synthesizing these principles, a robust, modern framework for drug repurposing can be proposed. This exemplary methodology integrates multi-modal features into a heterogeneous graph and leverages a Graph Attention Network (GAT) for prediction.

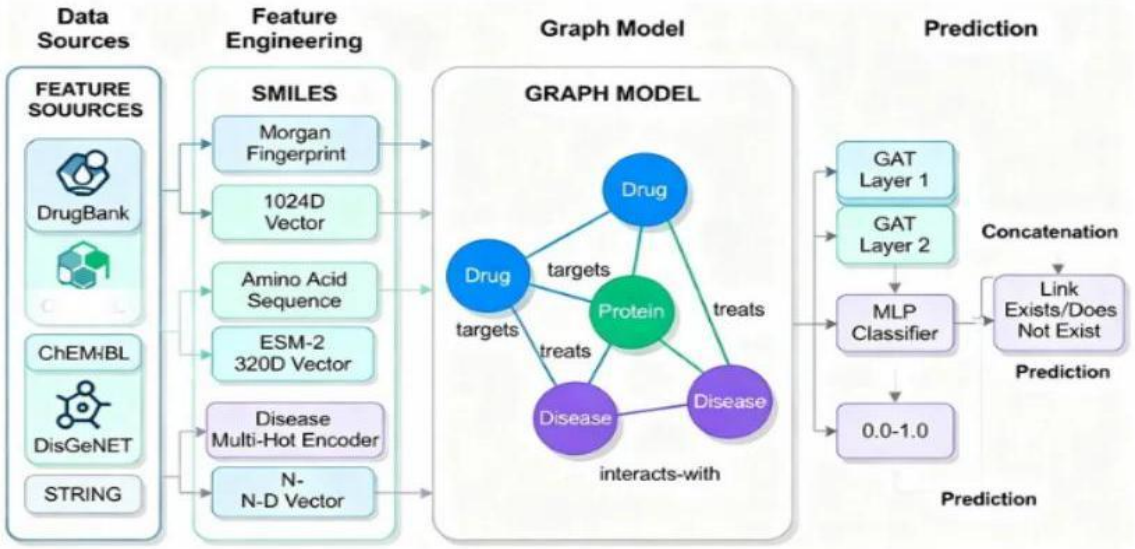

Fig. 1: Overview of the proposed GAT-based framework for drug repurposing.

Table 1: Summary of Referenced Works

Sr. No.	Authors	Year	Title	Methodology	Findings	Limitations	Evaluation Metric(s) and Score(s)
1	Otero-Carrasco et al.	2025	Prioritization of potential drugs through pathway-based drug repurposing and network proximity analysis	Pathway-based analysis combined with network proximity (Z-score) on a PPI network and transcriptomic data from CMap.	Integrated network proximity with gene expression to rank drug candidates; identified sorafenib as a promising candidate for Alzheimer's disease.	Relies on completeness and accuracy of public databases; CMap data from cancer cell lines may not represent all contexts.	Network proximity Z-score and CMap modZ; top candidate (Sorafenib for AD) had Zscore $= -4.4487$.
2	Otero-Carrasco et al.	2023	Orphan drugs and rare diseases: Unveiling biological patterns through drug repurposing	Descriptive and statistical analysis of orphan drug approvals, repurposing cases, and phenotypic similarity using Jaccard index.	Found that repurposing often occurs between two rare diseases; these pairs show higher phenotypic similarity.	Descriptive and pattern-based study, not predictive; limited rare disease data.	Jaccard Index $= 0.206$ (vs. 0.054 baseline); $p < 10e-26$.
3	Ceddia et al.	2020	Matrix factorization-based technique for drug repurposing predictions	Non-negative Matrix TriFactorization (NMTF) with shortest-path analysis on PPI to infer novel drug-protein interactions.	Enhanced predictions for proteins lacking known drug interactions.	Dependent on rank parameters; does not model biological pathway directionality.	APS $= 0.863$; AUC $= 0.931$.
4	Marín Tercero et al.	2024	Exploring drug repurposing opportunities for schizophrenia: A network medicine approach	Network proximity analysis to define a "disease module" for schizophrenia and compute drug-target distances.	Identified 14 potential repurposing candidates supported by literature.	Interactome incomplete; computational predictions need experimental validation.	Z-score -0.15 ; top drug Zscore $= -115.58$.
5	Castiglione et al.	2023	Explainable drug repurposing approach from biased random walks	Biased random walks on knowledge graphs modeled as ergodic Markov processes for explainable recommendations.	Provided accurate and efficient recommendations; promising rheumatoid arthritis candidates.	Sensitive to graph structure and completeness; sparse data may cause noise.	AUC (best) $= 0.96$; Accuracy $= 0.875$.

6	Ayuso-Munoz et al.	2023	Enhancing drug repurposing on graphs by integrating drug molecular structure as feature	GNN on heterogeneous graph with molecular structure-based initialization for drug nodes.	Molecular structure features improved GNN-based repurposing accuracy.	Performance depends on embedding quality; other node features excluded.	AUROC = 0.9148; AUPRC = 0.9219.
7	Park et al.	2025	Dual representation learning for predicting drug-side effect frequency using protein target information	Deep learning with dual embeddings for drugs and side effects in a common vector space.	Achieved SOTA prediction of side-effect frequencies, especially for unseen drugs.	Focused on side-effects, not direct repurposing; dependent on feature data.	AUROC = 0.901; AUPRC = 0.436.
8	Li and Hu	2024	Drug-target interaction prediction via deep multimodal graph and structural learning	Hybrid framework combining multimodal GNN and CNNbased structural learning.	Outperformed benchmarks; generalized well to unseen drugs/proteins.	High model complexity; optimal ensembling may be datadependent.	AUROC (DrugBank) = 0.973; AUPRC (DrugBank) = 0.954.
9	Rodriguez Gonzalez et al.	2025	DRIVE: A data-driven platform for disease visualization and drug repurposing	DRIVE integrates multiple data sources and six computational methods for repurposing.	Offers an interactive platform for exploring disease networks and hypotheses.	Platform paper; depends on integrated models for effectiveness.	Uses GNN prediction scores, proximity Z-scores; references prior benchmarks.
10	Artiñano-Muñoz et al.	2024	DRAGON: Drug repurposing via graph neural networks with drug and protein embeddings as features	GNN-based link prediction (DRAGON) using drug/protein embeddings from molecular and sequence data.	Multi-modal embeddings improved PR-AUC for drugdisease prediction.	Lacked embeddings for diseases/pathways; not validated experimentally.	PR-AUC = 0.945.
11	Bacciu et al.	2024	Deep graph networks for drug repurposing with multi-protein targets	Deep Graph Network predicting drug interactions with multiple protein targets jointly.	Modeling multi-protein targets improved robustness and prediction accuracy.	Performance depends on protein functional relatedness; computationally heavy.	AUROC (multi-protein) = 0.9413.
12	Wang et al.	2018	ANTENNA: A multi-rank, multi-layered recommender system for drug repurposing	Multi-layered recommender (ANTENNA) using trifactorization and Random Walk with Restart.	Predicted diazoxide as repurposed for TNBC, validated experimentally.	Complex model; performance sensitive to network density and data quality.	FDR = 0.0108 (Diazox-ide-TNBC); Recall@K benchmarked.

A. Data Preprocessing and Knowledge Graph Construction

The framework is an end-to-end pipeline that aims to turn raw, unconnected biomedical data into a ranked list of actionable hypotheses of drug repurposing. The first step is the curation and integration of data from several canonical databases to create a solid data foundation. Drug-focused data such as chemical structures (SMILES) and established protein targets are obtained from DrugBank and ChEMBL. Diseasegene associations are collected systematically from the DisGeNET database to establish a genomic ground for disease definitions. To simulate the underlying biological environment, a large PPI network is taken from the STRING database. A careful integration step is executed to align different entity identifiers to a uniform schema with precautions to ensure consistency of the dataset.

The integrated dataset is then structured as a large-scale, heterogeneous graph, formally defined as $G = (V, E, R, \phi, \psi)$, where V represents the set of all nodes (i.e., biological entities), and E represents the set of all edges (i.e., known relationships). R defines the set of all possible relation types. The function $\phi: V \to T_V$ maps each node to its specific type, where the set of node types is $T_V = \{$ drug, disease, protein $\}$. Similarly, the function $\psi: E \to R$ maps each edge to its corresponding relation type, such as treats, targets, or interacts with.

B. Multi-Modal Feature Generation Module

This module is responsible for creating the initial, highdimensional feature vectors that serve as input for our deep learning model. Each entity type is processed through a specialized embedding pipeline to capture its unique characteristics.

Drug Embedding: To capture the chemical and structural properties of drugs, Morgan fingerprints of size 1024 were generated from canonical SMILES strings using the RDKit library [13]. This algorithm creates a vector representing the presence or absence of specific circular substructures within a molecule, providing a rich numerical description of its topology that is crucial for predicting bioactivity.

Protein Embedding: To represent the functional context of proteins, we utilized ESM-2, a state-of-the-art protein language model [14]. The canonical amino acid sequence of each protein was fed into the model to generate a 320 dimensional embedding. This dense vector captures complex biochemical and evolutionary information learned from millions of sequences, providing a far richer representation than a simple categorical identifier.

Disease Embedding: To ground diseases in their underlying genomic basis, we created multi-hot encoded vectors using gene association data from DisGeNET [6]. Each vector's dimension corresponds to the total number of unique genes in our dataset, with a '1' indicating a known link between the disease and a specific gene. This creates a unique "genomic footprint" for each pathology.

C. Graph Representation Learning Module

This module is the core of our framework, designed to learn the complex, non-linear relationships within the knowledge graph. The initial feature vector for any node i in the graph is defined as $h_i^{(0)} = x_i$, where x_i is the feature vector generated in the previous module. The goal of the subsequent layers is to refine this initial representation into a final, context-aware embedding $h_i^{(L)}$ after L layers of graph convolution.

We employ a Graph Attention Network (GAT) which learns to weigh the importance of different neighbors during aggregation. The unnormalized attention score e_{ij} between a central node i and a neighbor j is calculated as:

$$e_{ij} = \text{LeakyReLU}\left(\mathbf{a}^T\left[\mathbf{W}h_i \| \mathbf{W}h_j\right]\right) \tag{7}$$

Here, h_i and h_j are the feature vectors, transformed by a learnable weight matrix $\mathbf{W}$ and concatenated. A dot product is taken with a learnable attention vector $\mathbf{a}^T$. These scores are normalized into attention weights α_{ij} using the softmax function:

$$\alpha_{ij} = \text{softmax}_j(e_{ij}) = \frac{\exp(e_{ij})}{\sum_{k \in \mathcal{N}_i} \exp(e_{ik})} \tag{8}$$

where $\mathcal{N}_i$ is the set of all one-hop neighbors of node i. The updated embedding for node i at the next layer is a weighted sum of its neighbors' transformed features:

$$h_i^{(l+1)} = \sigma\left(\sum_{j \in \mathcal{N}_i} \alpha_{ij} \mathbf{W} h_j^{(l)}\right) \tag{9}$$

By stacking two such layers, our model allows information to propagate across a two-hop neighborhood, enabling it to capture more complex, indirect relationships.

D. Prediction Module

After the GAT produces the final, context-aware embeddings $h^{(L)}$, the model must make a prediction. For a given drug-disease pair (v_d, v_p), their final embeddings are first concatenated. This combined vector is then passed through a simple Multi-Layer Perceptron (MLP), which acts as a binary classifier to produce a raw output score (logit) s_{dp} :

$$s_{dp} = \text{MLP}\left(\left[h_{v_d}^{(L)} \| h_{v_p}^{(L)}\right]\right) \tag{10}$$

This logit is converted into a probability using the sigmoid function, indicating the likelihood that drug v_d can be repurposed for disease v_p :

$$P(v_d, v_p) = \sigma(s_{dp}) = \frac{1}{1 + e^{-s_{dp}}} \tag{11}$$

IV. CHALLENGES AND FUTURE DIRECTIONS

Despite significant progress, several challenges remain in the field of computational drug repurposing.

Data Quality and Integration: The performance of all computational methods is contingent on the quality, completeness, and standardization of the underlying data. Integrating disparate data sources with varying identifiers and levels of evidence remains a significant hurdle.

Model Generalizability and Validation: Many models are trained on known associations and may not generalize well to novel drugs or diseases (the "cold start" problem). Furthermore, computational predictions require rigorous experimental validation, which is often a bottleneck. Standardized benchmarks and validation datasets are crucial for comparing methods robustly [9].

Explainability and Interpretability: As models, especially deep learning ones, become more complex, their "black-box" nature becomes a barrier to clinical translation. Methods that provide explainable predictions, such as tracing influential pathways or neighbors in a graph, are becoming increasingly important [11].

Biological Complexity: Current models often simplify complex biological realities. For instance, predictions are rarely tissue-specific, and most methods predict interactions with single targets rather than multi-target complexes or pathways [3].

Future work will likely focus on addressing these challenges. We anticipate a greater emphasis on multi-modal learning, where models can seamlessly integrate even more diverse data types, including electronic health records, imaging data, and transcriptomics. There is a growing need for models that can handle dynamic networks to capture changes over time or in different cellular contexts. Finally, the push for explainable AI (XAI) will continue, leading to hybrid models that combine the predictive power of deep learning with the interpretability of network-based approaches.

V. CONCLUSION

This survey has provided a comprehensive overview of the modern computational landscape for drug repurposing, charting its evolution from serendipitous discoveries to systematic, data-driven science. We have detailed the principal methodologies that are currently shaping the field, beginning with foundational network medicine concepts like pathway and proximity analysis, which leverage the topological structure of biological networks to infer therapeutic relationships. The discussion moved to machine learning techniques, including matrix factorization methods that treat repurposing as a recommendation problem, and culminated in the exploration of advanced deep learning models. In particular, we highlighted the impact of Graph Neural Networks (GNNs), which are uniquely suited to learning from the complex, interconnected nature of biomedical data. A recurring theme has been the critical importance of integrating heterogeneous, multi-modal data and developing rich feature representations for drugs, proteins, and diseases. While these computational frameworks offer immense potential to accelerate therapeutic discovery, we also acknowledged the persistent challenges of data integration, model generalizability, and the crucial need for explainability to bridge the gap between in silico predictions and clinical translation. Future progress will undoubtedly depend on the development of more sophisticated, interpretable, and biologically-informed models that can navigate the complexity of human disease and pharmacology.

REFERENCES

[1] B. Otero-Carrasco et al., “Orphan drugs and rare diseases: Unveiling biological patterns through drug repurposing,” Proc. IEEE 36th Int. Symp. Comput.-Based Med. Syst. (CBMS), 2023, pp. 123–130.
[2] G. Ceddia et al., “Matrix factorization-based technique for drug repurposing predictions,” IEEE J. Biomed. Health Inform., vol. 24, no. 11, pp. 3100–3110, 2020.
[3] D. Bacciu et al., “Deep graph networks for drug repurposing with multiprotein targets,” IEEE Trans. Emerg. Top. Comput., vol. 12, no. 1, pp. 177–188, 2024.
[4] A. Wang et al., “ANTENNA, a multi-rank, multi-layered recommender system for inferring reliable drug-gene-disease associations,” IEEE/ACM Trans. Comput. Biol. Bioinform., vol. 15, no. 6, pp. 1960–1967, 2018.
[5] H. Li and Z. Hu, “Drug-target interaction prediction via deep multimodal graph and structural learning,” Proc. IEEE Eng. Med. Biol. Soc. (EMBC), 2024, pp. 1–6.
[6] J. Pinero ˜ et al., “The DisGeNET knowledge platform for disease genomics: 2019 update,” Nucleic Acids Res., vol. 48, no. D1, pp. D845–D855, 2020.
[7] M. Mar´ın Tercero et al., “Exploring drug repurposing opportunities for schizophrenia: A network medicine approach,” Proc. IEEE 37th Int. Symp. Comput.-Based Med. Syst. (CBMS), 2024, pp. 101–108.
[8] A. Rodriguez Gonzalez et al., “DRIVE: A data-driven platform for disease visualization and drug repurposing,” Proc. IEEE 38th Int. Symp. Comput.-Based Med. Syst. (CBMS), 2025, pp. 504–505.
[9] A. Otero-Carrasco et al., “Prioritization of potential drugs through pathway-based drug repurposing and network proximity analysis,” Proc. IEEE 38th Int. Symp. Comput.-Based Med. Syst. (CBMS), 2025, pp. 246–251.
[10] F. Castiglione et al., “Explainable drug repurposing approach from biased random walks,” IEEE/ACM Trans. Comput. Biol. Bioinform., vol. 20, no. 2, pp. 1009–1018, 2023.
[11] R. Artinano-Mu ˜ noz ˜ et al., “DRAGON: Drug repurposing via graph neural networks with drug and protein embeddings as features,” Proc. IEEE 37th Int. Symp. Comput.-Based Med. Syst. (CBMS), 2024, pp. 170–175.
[12] A. Ayuso-Munoz et al., “Enhancing drug repurposing on graphs by integrating drug molecular structure as feature,” Proc. IEEE 36th Int. Symp. Comput.-Based Med. Syst. (CBMS), 2023, pp. 192–197.
[13] S. Park et al., “Dual representation learning for predicting drug-side effect frequency using protein target information,” IEEE J. Biomed. Health Inform., vol. 29, no. 3, pp. 1817–1827, 2025.
[14] G. Landrum et al., RDKit: Open-source cheminformatics, 2006. [Online]. Available: https://www.rdkit.org
[15] D. Rogers and M. Hahn, “Extended-connectivity fingerprints,” J. Chem. Inf. Model., vol. 50, no. 5, pp. 742–754, 2010.
[16] Z. Lin et al., “Evolutionary-scale prediction of atomic-level protein structure with a language model,” Science, vol. 379, no. 6637, pp. 1123–1130, 2023.