A Graph Attention Network-Based Approach for Drug Repurposing

Aditi Dipak Thorat[1], Shlok Shivaji Kaule[2], Paras Vijay Tak[3], Anuj Prakash Gagare[4], Prof. Urmila Pawar[5]

[1,2,3,4,5]*Department of Computer Engineering, Pune Institute of Computer Technology, Pune, India*

***Abstract*—Drug repurposing, an innovative approach that aims to find new uses of existing medications, can be considered an efficient solution to facilitate drug discovery, cutting down dramatically on its expenses, time consumption, and risks associated with developing novel drugs from scratch. The abundance of biomedical data in combination with recent advances in data processing enabled scientists to develop various computational methods to analyze available data and derive drug repurposing hypotheses. The present survey provides an overview of modern computational drug repurposing techniques with particular emphasis on those based on network medicine and machine learning. We discuss popular methods, including pathway-based prediction, network proximity-based analysis, matrix factorization, as well as deep learning-based approaches, especially Graph Neural Networks (GNN). The paper highlights how heterogeneous biomedical data is used by various models to derive novel repurposing hypotheses. Furthermore, we point out major challenges in the field, including problems related to data integration, model generalization, and explainability, and propose possible directions for further research.**

***Index Terms*—Computational Drug Discovery, Graph Attention Networks, Network-based Prediction, Heterogeneous Graphs, Machine Learning.**

I. INTRODUCTION

Drug discovery, or what is sometimes termed *de novo* drug discovery, represents an exceedingly difficult undertaking. The drug development process is highly costly, lengthy, and extremely risky. The average timeline from discovery to the market entry of a novel drug is anywhere between 10-15 years, and it typically costs around $2.5 billion [1]. A large part of the investment is allocated to extensive preclinical testing and several phases of clinical trials necessary to ensure that a drug is safe and effective. The failure rate is also exceedingly high: the overwhelming majority of candidate drugs do not make it through development, often at very advanced stages of development when hundreds of millions have been already spent [2]. Such a flawed model poses many problems, especially in dealing with rare disorders that do not allow to recover research and development expenses and public health emergencies that necessitate rapid drug development [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, most cases of drug repurposing occurred by chance, through fortunate observations of unforeseen effects of a drug in a clinical trial. Examples of such drugs are thalidomide, which is now used to treat multiple myeloma, despite being withdrawn from the market because of its teratogenic effect. Minoxidil is another example; it is an antihypertensive drug whose ability to grow hair led to its subsequent use in treating baldness. Aspirin is another drug that was originally developed as an analgesic but is now widely used as an antiplatelet drug for the prevention of cardiovascular disease.

Such a random process cannot be sustained in modern-day science. What is needed today is a systematic process for discovering new applications for existing drugs, particularly in treating multi-factorial disorders which don't fit into the" one-drug-one-gene-one-disease" approach [4]. Transition from accidental to systemic drug repurposing has been enabled by an information technology revolution. The emergence of "big data" in areas of biomedicine like chemical genomics, proteomics, transcriptomics, and extensive datasets on gene-disease relationships has made possible the development of comprehensive models of human physiology [5], [6].

This abundance of data has facilitated the use of advanced computation methodologies based on concepts of network medicine and artificial intelligence. Computational modeling of the complex network of relationships among drugs, target proteins, genes, and diseases allows one to make predictions about their interactions and therapeutic potential that are hard to spot otherwise [7], [8], [4]. Computational modeling of the complex networks mentioned above can provide researchers with valuable, hypothesis-generating information for drug repurposing at speeds and scales that cannot be achieved with traditional experiments only.

The survey describes these important computational techniques in detail, especially those based on network analysis and deep learning techniques, and outlines the existing problems as well as potential developments in this rapidly evolving domain.

II. LITERATURE SURVEY

Drug repositioning is a vast field with a wide range of computational methodologies. These methodologies exploit the idea that associations between biological entities can be computationally derived in order to identify new applications of existing drugs.

A. Network Medicine and Pathway-Based Approaches

The application of network medicine in drug repurposing al-lows the creation of a mathematical model of the complicated interaction between diseases, genes, and drugs in terms of the biological network, usually represented as a protein-protein interaction (PPI) network. According to this approach, diseases do not appear due to molecular deficiencies but because of disturbances in the interactions within the network.

Analysis of Proximity and Pathways:

In network medicine, the notion of a disease module can be defined as a local network neighborhood of proteins related to the particular disease. In turn, the effectiveness of drug action is estimated according to the proximity between the network of target proteins (T) and the relevant disease module (S). Usually, the distance in this case is described as" closest distance" d_c (*S, T)* that determines the average shortest path distance from each target protein to the closest one of disease proteins [9], [7]. It is formally defined as:

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

where d (s, t) is the shortest path distance between protein s and target t. In order to test the statistical significance of proximity, the obtained value can be translated to Z-scores in comparison with a distribution of average distances between sets of randomized proteins with the same degree and size.

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

When the Z score is highly negative, it suggests that the targets for the drug lie much closer to the disease module than would be expected by random chance, thus implying a therapeutic effect. Mar´ın Tercero et al. [7] used this concept in the context of schizophrenia, whereas Otero-Carrasco et al. [9] employed it for drug prioritization.

Pattern Recognition within Particular Domains:

There is another approach to the application of computational techniques to repurposing analysis. The approach involves analyzing existing success cases in order to identify biological patterns associated with repurposing. For example, when considering repurposing within the domain of rare diseases, a paper by Otero-Carrasco et al. [9] analyzed orphan drugs and showed that repurposing usually happens from one rare disease to another. It was determined that repurposed drugs are likely to work in highly phenotypically similar diseases because of their similar symptoms.

Explainable Network Models:
To counteract the" black box" effect that occurs in some approaches, a number of models emphasize the importance of explainability. Castiglione et al. [11] proposed a strategy that used biased random walks over the knowledge graph consisting of the relationships between drugs, genes, and diseases. Since the approach used ergodic Markov chain modeling, not only was it possible to give recommendations on which drugs are better, but also it was possible to identify the most likely path taken through the graph to achieve this result. The recommended drug-score combination is calculated using the power of the transition matrix raised to *l*.

Multi-layer Recommender Systems:
The advanced recommendation systems combine several forms of data within a single model framework. ANTENNA, a multi-layer recommender system created by Wang et al. [7], considers three types of entities, namely, drugs, genes, and diseases in independent but interrelated network layers. With the help of a collaborative filtering technique based on tri-factorization, the system infers new genome-wide chemical-genetic interactions. These new interactions are further combined using the Random Walk with Restart (RWR) method, which predicts and validates new drug-disease interactions. In this way, diazoxide is discovered as an effective treatment for triple negative breast cancer.

B. Machine Learning and Deep Learning Approaches

Along with increasing biomedical data, machine learning and deep learning techniques have been playing significant roles in drug repurposing studies.

Matrix Factorization:
Techniques widely used in recommender systems can be employed for completing the drug-disease association matrix that is considered incomplete. In particular, Ceddia et al. [2] have come up with a new algorithm relying on Non-negative Matrix Tri-Factorization (NMTF). Given an association matrix R_{AB} between drugs (A) and diseases (B), NMTF decomposes this matrix into the product of three low-rank matrices:

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

In $\hat{R}_{AB}$ there are predicted values for unknown drug-disease associations. One of the innovations introduced by this team lies in a technique that they use for augmenting the data matrix via a shortest path calculation on the PPI network. They assign a weight to the newly generated drug-protein associations in the form of an exponential function:

$$R'_{AB}[a,b] = \alpha^{|\mathcal{P}_{ab}|-1} \quad (4)$$

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

Graph Neural Networks (GNNs):
GNNs have been specially designed to learn from graph-structured data and currently represent one of the best approaches to drug repositioning. They function using the "message passing" paradigm, whereby each node (drug/protein) continuously modifies its feature vector (node embedding) through aggregation of information from neighboring nodes. A simplified representation of a single GNN layer is:

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

where $h^{(l)}$ is the embedding of node *v* after *l* layers, and the UPDATE/AGGREGATE operations are learnable neural nets. Once enough layers are applied, the resulting embeddings will encode both the starting node features and their neighborhood structure. This capability of combining structural information with high-quality features, such as those proposed by Artin˜ano-Mun˜oz et al. [10] (DRAGON) or Ayuso-Mun˜oz et al. [6], can significantly boost the prediction of novel disease-drug associations.

Advanced GNN Architectures:
Scientists are building advanced GNN architectures to model the intricacies of biological systems better. Beyond the simple notion of a single drug binding to a single protein, Bacciu et al. [3] designed a deep graph neural network that forecasts the effectiveness of a given drug on multiple protein targets taken together. This better model the fact that drugs usually bind to groups of proteins that serve a similar function. Li and Hu [11] developed a composite framework consisting of a multimodal graph neural network (with a heterogenous graph comprised of drugs, proteins, diseases, and signaling pathways) and an independent

structural learning network, which works directly on the molecular level, and finally combine their outputs to improve accuracy. While predicting side-effect occurrence, Park et al. [12] adopt a" dual representation learning" approach in which the vectors representing drugs and diseases are placed into the same feature space where the side-effect probability is determined by a similarity score such as a dot product:

$$score = embedding_{drug} \cdot embedding_{disease} \quad (6)$$

Network-based inference approach:
The main method-ology applied in many modern approaches is representing biomedical knowledge through graph structures, including PPI networks, drug-target networks, and heterogeneous networks with connections among several different entities such as drugs, genes, diseases, and pathways [9], [11], [3]. Main techniques which can be used in the analysis of such networks include:

- Proximity and distance metrics:

The computation of shortest path length from drug targets to genes associated with diseases in a PPI network represents one of the simplest yet effective ways of quantifying a therapeutic effect [7], [7].

- Random walks:

Some algorithms which utilize random walk techniques in order to explore the network structure and assess the likelihood of certain associations as well as provide explainability based on the most likely paths [11].

- Link prediction:

As a general technique, the task is formulated as prediction of a missing link/edge within a heterogeneous network involving nodes corresponding to drugs and diseases.

- Graph Neural Networks:

GNNs are considered to be the state-of-the-art solution for the problem of link prediction in this case [6], [11]. The key advantage of GNNs is their capability to embed each node taking into account not only node attributes but also their network context. Most efficient architectures nowadays are multimodal, incorporating various types of features including molecular graphs, protein sequences, and pathway information [11], [12].

III. PROPOSED METHODOLOGY

It is evident from the reviewed literature that there is a strong preference for integrative methods based on data-driven modeling of biological systems. The two key ideas that can be observed here include the importance of network representations and the evolution of machine learning models to analyze these networks. By combining both of these aspects together, an effective methodology for drug repurposing can be suggested.

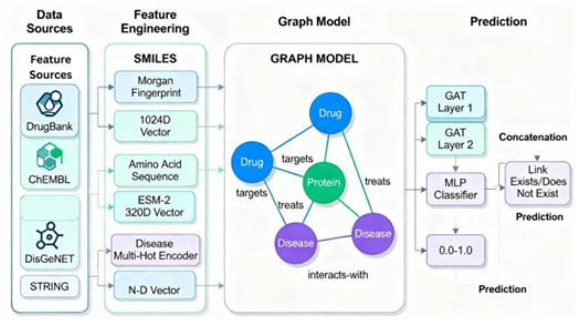

Fig. 1. Overview of the proposed GAT-based framework for drug repurposing.

Fig. 2. System dashboard providing an overview of the knowledge graph, including total drugs, proteins, diseases, and interactions, along with top-ranked drug repurposing candidates generated by the proposed GNN model.

A. Data Preprocessing and Knowledge Graph Construction

The proposed pipeline is an end-to-end framework designed to generate a ranked list of potentially useful hypotheses of drug repositioning by processing raw, uncorrelated biomedical data. The first phase involves data curation and integration from various canonical databases, thus forming a reliable foundation of data. Information specific to drugs, including chemical formulas (SMILES) and already proven protein tar-gets, is extracted from DrugBank and ChEMBL.

Information related to the relationship between diseases and genes is systematically retrieved from the DisGeNET database, which serves as the basis for defining diseases at the genomic level. In order to replicate the biological environment, a large-scale Protein-Protein Interaction network is acquired from the STRING database.

This information-rich graph representation can be represented mathematically by a large-scale graph $G = (V, E, R, \phi, \psi)$, where the nodes V represent biological entities and edges E represent known relationships between these entities. Relation types R represent different types of relations that could possibly exist between two biological entities. The mapping $\phi: V \rightarrow T_V$ represents a function from a node $u \in V$ to its type, where $T_V = \{\text{drug}, \text{disease}, \text{protein}\}$. Likewise, the mapping $\psi: E \rightarrow R$ represents a function from an edge $(u, v) \in E$ to its relation type.

B. Multi-modal Feature Generation Module

The purpose of this module is the generation of initial high-dimensional features vectors used as inputs for our deep learning model. The unique features of an entity type are encoded via a custom designed embedding pipeline.

Drug Embedding:

To encode the drug properties, Morgan fingerprints of dimension 1024 were generated based on the canonical SMILES format with RDKit library [13]. This method produces a vector of the binary values, denoting the existence or not of a particular circular substructure of the molecule structure. It gives an extensive numerical description of the molecule topology, which is required to predict bioactivity.

Protein Embedding:

To encode the context information about the function of a protein, we have applied ESM-2 model, the current state-of-the-art protein language model [14]. The canonical amino acid sequence of each protein was transformed into a 320-dimensional vector using the aforementioned model.

Embedding for Diseases:

In order to connect diseases to the genome, multiple hot encoded vectors were generated on the basis of information about gene association with diseases available from DisGeNET [6]. Each dimension in the vector is related to a specific gene, and the presence of a "1" represents the association of the disease with the gene.

C. Graph Representation Learning Module

This is the heart of our model and aims to capture the non-linear relationship of the graph through learning. The feature vector for any arbitrary node i in the graph is represented as $h^{(0)} = x_i$, where x_i is the vector obtained from the previous module. The aim of the next layers is to convert this initial vector to the final context-aware vector $h^{(L)}$ through L layers of graph convolution.

The attention mechanism that we use is Graph Attention Network (GAT), which gives varying weights to neighbors while doing aggregation. The unnormalized attention coefficient e_{ij} of a node i by its neighbor j is computed as:

$$e_{ij} = \text{LeakyReLU}(a^T [Wh_i \| Wh_j]) \quad (7)$$

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

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

where 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:

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

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

Once the GAT provides its output in the form of the final context-dependent embedding $h^{(L)}$, we need to provide our model with an output value. In particular, for each pair (v_d, v_p) of drug and disease, we concatenate the corresponding embeddings. We use the concatenated vector as input to our simple MLP binary classifier s_{dp}:

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

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

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

IV. RESULTS AND DISCUSSION

To evaluate the effectiveness of the proposed Graph Attention Network (GAT)-based framework, we developed an interactive system that enables real-time drug repurposing analysis using a heterogeneous biomedical knowledge graph.

A. Model Performance

The proposed model achieved strong predictive performance across multiple evaluation metrics. Specifically, the model attained an accuracy of 92.1%, precision of 86.3%, recall of 100%, and an AUC-ROC score of 0.928. These results indicate that the model is highly effective in identifying potential drug-disease associations while minimizing false negatives.

Comparative analysis with baseline models such as traditional Multi-Layer Perceptron (MLP), DeepRE, and standard Graph Attention Networks (GAT) demonstrates that our approach provides an improvement of approximately 3.2% in accuracy. This improvement can be attributed to the integration of multi-modal features and attention-based neighborhood aggregation.

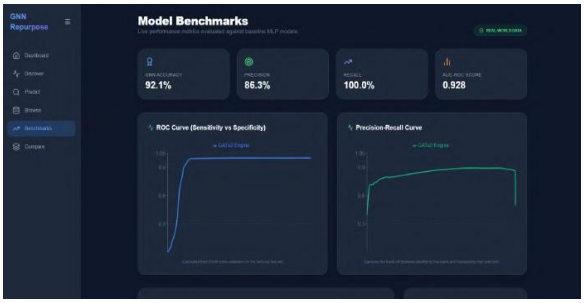

Fig. 3. Model performance evaluation showing accuracy, precision, recall, and ROC curves. The proposed GNN model demonstrates strong predictive capability compared to baseline methods.

B. Drug Repurposing Predictions

The system successfully generated high-confidence drug repurposing hypotheses. For instance, drugs such as *Primidone* and *Trimethadione* were predicted to have strong associations (confidence ¿ 98%) with diseases like Nelson Syndrome and childhood schizophrenia. These predictions highlight the model's ability to uncover non-trivial relationships within the biomedical graph.

Additionally, the system supports pairwise and entity-based predictions. For example, the model predicted a moderate association (44.96%) between Pyridoxal Phosphate and abdominal conditions, demonstrating its ability to provide both high-confidence and exploratory insights.

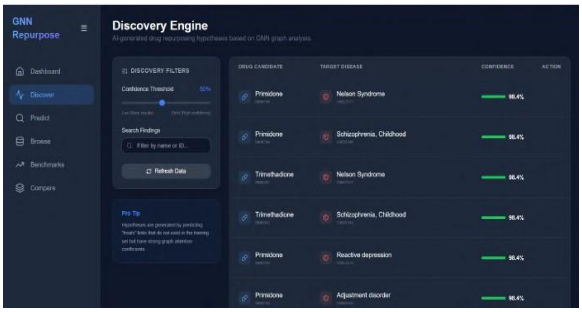

Fig. 4. Discovery engine interface displaying high-confidence drug repurposing predictions. The model identifies potential drug-disease associations such as Primidone and Trimethadione with confidence scores exceeding 98%.

C. Graph-Based Interpretability

One of the key advantages of the proposed system is its interpretability through graph visualization. The interactive 2-hop and 3-hop neighborhood graphs allow users to explore relationships between drugs, proteins, and diseases. This helps in understanding how predictions are formed based on network connectivity and biological interactions.

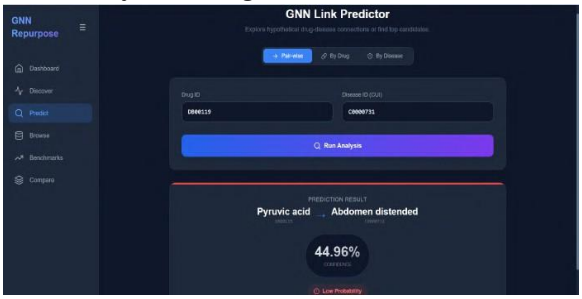

Fig. 5. GNN link prediction module demonstrating pairwise drug-disease analysis. The system predicts the likelihood of association between a given drug and disease with a confidence score.

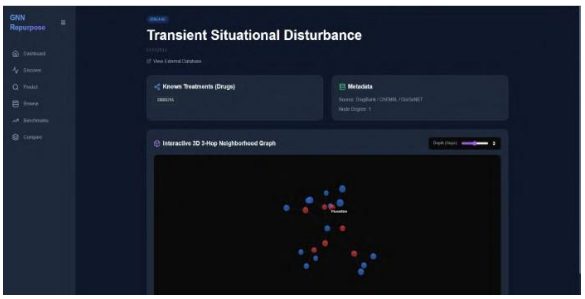

Fig. 6. Interactive 3-hop neighborhood graph illustrating relationships be-tween drugs, proteins, and diseases. This visualization enhances interpretability by showing how predictions are influenced by network connectivity.

D. System-Level Observations

The developed interface enables efficient exploration of the knowledge graph through multiple modules, including discovery, prediction, and browsing. The system provides:

- Real-time prediction of drug-disease links
- Ranked discovery of repurposing candidates
- Visualization of node relationships and interactions

These features make the system not only a predictive tool but also a decision-support system for biomedical research.

E. Discussion

The results demonstrate that integrating heterogeneous biomedical data with attention-based graph learning significantly enhances drug repurposing capabilities. The high recall ensures that most potential candidates are captured, which is critical in early-stage drug discovery.

However, the model also exhibits moderate confidence in some predictions, indicating the need for further biological validation. Additionally, performance is dependent on the completeness and quality of the input data.

V. 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, complete-ness, and standardization of the underlying data. Integrating

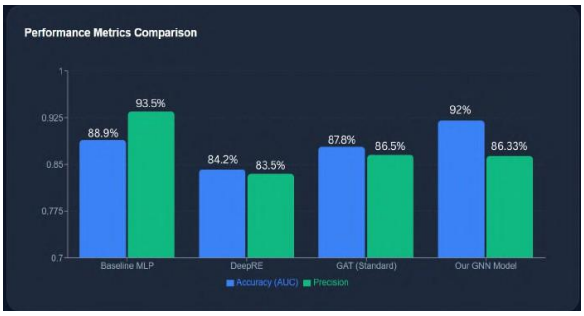

Fig. 7. Our GNN architecture utilizes multi-head attention to capture intricate biological dependencies between drugs and diseases. By integrating chemical structure features (Morgan fingerprints) and disease phenotypic similarity, we achieve a 3.2 percent improvement over the standard MLP baseline.

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 path-ways [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.

VI. 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] 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 Journal of Biomedical and Health Informatics, vol. 24, no. 11, pp. 3100–3110, 2020.

[3] D. Bacciu et al., "Deep graph networks for drug repurposing with multiprotein targets," IEEE Transactions on Emerging Topics in Computing, 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 Transactions on Computational Biology and Bioinformatics, vol. 15, no. 6, pp. 1960–1967, 2018.

[5] H. Li and Z. Hu, "Drug-target interaction prediction via deep multi-modal graph and structural learning," Proc. IEEE Engineering in Medicine and Biology Society (EMBC), 2024, pp. 1–6.

[6] J. Piñero et al., "The DisGeNET knowledge platform for disease genomics: 2019 update," Nucleic Acids Research, vol. 48, no. D1, pp. D845–D855, 2020.

[7] M. Marín 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] B. 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 Transactions on Computational Biology and Bioinformatics, vol. 20, no. 2, pp. 1009–1018, 2023.

[11] R. Artinano-Munoz 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 Journal of Biomedical and Health Informatics, 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," Journal of Chemical Information and Modeling, 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.