©2025 IEEE. Personal use of this material is permitted. Permission from IEEE must be obtained for all other uses, in any current or future media, including reprinting/republishing this material for advertising or promotional purposes, creating new collective works, for resale or redistribution to servers or lists, or reuse of any copyrighted component of this work in other works.

Explainable AI for Knowledge Graph-Based Drug Repurposing: Methods, Challenges, and Interpretability Frameworks

Interdisciplinary AI innovations in business

Marwa Mustafa
Engineering and Information Technology
University of Technology Sydney
Sydney, Australia
marwa.r.mustafa@student.uts.edu.au
0009-0009-5346-6254

Firas Al-Doghman
Engineering and Information Technology
University of Technology Sydney
Sydney, Australia
Firas.Al-Doghman@uts.edu.au
0000-0002-1020-8097

Farookh Hussain
Engineering and Information Technology
University of Technology Sydney
Sydney, Australia
Farookh.Hussain@uts.edu.au
0000-0003-1513-8072

Mohamed Awadallah
Engineering and Information Technology
University of Technology Sydney
Sydney, Australia
Mohamed.Awadallah@uts.edu.au
0000-0002-2317-4527

Mohammed Ali
School of Engineering
DSTI School of Engineering
Paris, France
mohammed.ali@edu.dsti.institute
0000-0001-8907-2374

***Abstract* - The increasing reliance on Artificial Intelligence (AI) in drug repurposing has introduced interpretability concerns, particularly when using complex models over biomedical Knowledge Graphs (KGs). This paper critically investigates the role of Explainable Artificial Intelligence (XAI) in enhancing the transparency and interpretability of AI-driven predictions in KG-based drug repurposing. This paper presents a structured taxonomy of XAI approaches - including graph-based feature attribution, path-based reasoning, counterfactuals, and rule-based logic - tailored for knowledge-rich biomedical graphs. Furthermore, an analysis is conducted on current limitations in XAI evaluation, the tension between explainability and predictive accuracy, and the need for audience-specific interpretability frameworks in biomedical contexts. By positioning XAI as a bridge between opaque machine learning models and clinically relevant reasoning, this paper contributes to the development of trustworthy AI systems in biomedical knowledge environments.**

Keywords - Explainable AI; Knowledge Graphs; Drug Repurposing; Graph Neural Networks, Link Prediction; Biomedical AI;

I. INTRODUCTION

The development and commercialization of new pharmaceutical agents is an inherently complex, resource-intensive endeavour. Empirical analyses estimate the average timeline for bringing a novel drug to market at approximately 8.3 ± 2.8 years, with associated expenditures beginning at $374 million and potentially exceeding $1.3 billion when accounting for high attrition rates. Notably, fewer than 10% of drug candidates ultimately achieve regulatory approval. This low success rate is largely attributable to the multi-phase clinical trial process mandated to ensure safety and efficacy. Phase I trials assess toxicity profiles and tolerability, with success rates varying between 3.4% and 32.6%. Phase II trials focus on identifying optimal dosing regimens and evaluating therapeutic potential, yet nearly 60% of compounds are discontinued at this stage due to inadequate efficacy or unacceptable adverse effects. Phase III trials are designed to confirm clinical effectiveness and establish a risk benefit profile suitable for market authorization. The cumulative duration from initial human testing to regulatory approval often spans a decade or more, imposing significant temporal and financial constraints on drug innovation pipelines [1] [2] [3] [4] [5].

The substantial financial burden associated with de novo drug development has led pharmaceutical companies to concentrate their efforts on conditions with large and profitable patient populations, often to the detriment of rare or neglected diseases. This market-driven research strategy contributes to a persistent innovation gap in areas with limited commercial incentive. In response, drug repurposing - also referred to as drug repositioning - has gained traction as a pragmatic alternative. This strategy seeks to identify novel therapeutic indications for existing pharmacological compounds, including those that failed to achieve regulatory approval for their original targets but possess favourable safety profiles or mechanistic relevance in other disease domains. By leveraging pre-existing pharmacokinetic and toxicological data, drug repurposing offers the potential to accelerate development timelines, substantially reduce costs, and mitigate the risk of clinical attrition, thereby providing a more efficient route to address unmet medical needs [2] [3].

The COVID-19 pandemic underscored critical vulnerabilities in global healthcare systems, particularly in the timely development and deployment of effective therapeutic interventions. Amid this public health crisis, the strategic re-evaluation of existing pharmacological agents - commonly referred to as drug repurposing - emerged as a vital response to accelerate treatment discovery. By circumventing the extensive safety validation phases required for novel compounds, repurposing strategies offer a time-efficient and cost-effective alternative to traditional drug development pipelines [5].

Recent advances in computational biology have markedly enhanced the feasibility and scalability of drug repurposing. In particular, Knowledge Graphs (KGs) have become essential tools for structuring heterogeneous biomedical data

XXX-X-XXXX-XXXX-X/XX/$XX.00 ©20XX IEEE

into semantically rich networks of entities (e.g., drugs, genes, diseases) and relationships. When combined with Artificial Intelligence (AI) and Machine Learning (ML) techniques, KGs facilitate the systematic identification of novel therapeutic hypotheses through link prediction, feature learning, and automated reasoning [1] [3].

AI-driven analysis enables the efficient extraction of complex relational patterns from large-scale biomedical datasets, revealing non-obvious connections that may inform repurposing opportunities. Nevertheless, existing KG-based systems often exhibit limited support for real-time data integration and lack transparent decision-making mechanisms - barriers that constrain their clinical utility and regulatory acceptance [3].

Despite these challenges, KGs provide a robust framework for integrating and navigating multi-source biomedical knowledge. Their capacity to unify diverse data modalities and elucidate latent associations renders them increasingly central to modern drug discovery and translational medicine. The following sections explore the foundational principles of KGs, delineate their structure and functionality, and examine their application in drug repurposing contexts [1].

This study hypothesizes that post-hoc explainability methods applied to biomedical knowledge graphs can generate stakeholder-relevant, biologically plausible explanations for drug repurposing predictions.

II. XAI IN KGS-BASED DRUG REPURPOSING

XAI aims to make AI models more transparent, interpretable, and accountable by providing human-understandable explanations of their decision-making processes. The concept was first formally defined in 2004 as the ability of an AI system to explain its actions. A widely accepted definition states: "Given an audience, an XAI is one that produces details or reasons to make its functioning clear or easy to understand" [6] [7].

Clarifying the distinction between interpretability and explainability is crucial in XAI. Interpretability refers to how easily a human can understand a model's internal mechanics-essentially, it's about the model's transparency. In contrast, explainability pertains to methods or processes that elucidate or justify a model's decisions to users [8].

AI models, especially deep learning-based ones, often function as "black boxes" where inputs and outputs are known, but the internal decision-making process is not transparent. This lack of interpretability creates significant challenges in trust, accountability, and regulatory compliance, particularly in sensitive domains like healthcare.

XAI is essential in the field of drug repurposing, as it enhances the transparency and trustworthiness of AI-driven predictions. One of its key advantages is providing clear explanations of why a particular drug is suggested for a new therapeutic use, helping researchers and clinicians validate AI-generated hypotheses. This is particularly critical in biomedical applications, where inaccurate predictions can have serious implications for patient safety. Additionally, Certain XAI methods can aid in uncovering biases or anomalous decision patterns in predictive models, though this depends on the granularity and fidelity of the explanations generated, ensuring that drug repurposing methods are scientifically sound and based on robust, interpretable evidence rather than black-box assumptions [9] [10] [11].

Interpretable Machine Learning (IML), Explainable AI (XAI), and Comprehensible AI (CAI)

While traditional approaches to explainability predominantly focus on XAI, which employs post-hoc interpretation techniques to clarify the decisions of otherwise opaque ("black-box") models, recent research highlights the complementary role of Interpretable Machine Learning (IML)-methods designed to produce models that are inherently understandable without additional explanation mechanisms. To bridge these two paradigms, Comprehensible Artificial Intelligence (CAI) has emerged as a unified framework that encapsulates both XAI and IML. By integrating post-hoc explanations and intrinsic interpretability within a structured approach, CAI supports the development of AI systems that are not only accurate but also inherently transparent. This combined perspective effectively addresses the critical requirements of trust, transparency, and accountability essential for real-world AI deployments.

A. XAI Types

XAI methods are broadly classified into the following types xAI Survey:

- Transparent box design: models designed to be interpretable by nature or by themselves, such as Decision Trees, Logistic Regression, general additive models, and Bayesian models.
- Post-hoc Explainability: refers to methods used after training a model to understand its decisions. Some common algorithms include SHAP (Shapley Additive Explanations), LIME (Local Interpretable Model-agnostic Explanations), Grad-CAM (Gradient-weighted Class Activation Mapping), Integrated Gradients, Layer-wise Relevance Propagation(LRP), and Counterfactual Explanations [12] .

B. Role of KG in XAI

KGs play a crucial role in enhancing AI interpretability by providing a structured representation of knowledge. Unlike black-box AI models, which lack transparency, KGs facilitate structured reasoning, rule induction, and logical inference, thereby supporting the generation of explanations that are interpretable by domain experts. These characteristics render KGs particularly suitable for domains where transparency and traceability are essential, such as healthcare and drug repurposing, where AI-generated predictions must be scientifically justified.

In the context of KGs, knowledge representation can be categorized into three paradigms: symbolic, sub-symbolic, and neuro-symbolic. Symbolic representations explicitly encode knowledge through logical triples and structured semantics, making them inherently interpretable and transparent. Sub-symbolic representations, on the other hand, encode knowledge implicitly through numerical embeddings, which, although effective in capturing latent patterns in data, typically lack interpretability due to their non-symbolic representations. Neuro-symbolic representations integrate symbolic logic-based structures and sub-symbolic embeddings, combining the interpretability of symbolic methods with the predictive power of sub-symbolic approaches. This hybrid form has become increasingly relevant for XAI, as it provides both accurate predictions and meaningful, human-understandable explanations, crucial for

building trustworthy and accountable AI systems. A recent work in drug repurposing demonstrates the potential of neurosymbolic learning to improve both accuracy and explainability. The study proposes a hybrid model that integrates KGs embeddings with symbolic biomedical rules, enabling drug-disease predictions that are not only data-driven but also interpretable.

TABLE I. TAXONOMY OF EXPLAINABLE AI METHODS FOR KNOWLEDGE GRAPH-BASED DRUG REPURPOSING

XAI Category	Method	Model Type	Explanation Output	Strengths	Limitations	References
Graph-Based Attribution	GNNExplainer	GNN	Minimal subgraph	Localized, model-agnostic	Sensitive to noise	[13]
	GraphLIME	GNN	Linear weights over features	Local interpretability	High computational cost	[14]
	GraphSHAP	GNN, KGE	SHAP values	Theoretically grounded fairness	Complexity in sampling	[15]
Path-Based	eXpath	Embedding models	Ontological paths	Semantically rich, symbolic	Limited to known paths	[16]
	Power-Link	GNN	Multi-hop path scores	Scalable, efficient	May miss long-range relations	[14]
Surrogate Models	PGM-Explainer	GNN	Bayesian network	Probabilistic uncertainty handling	Requires distribution estimation	[17]
	GraphSVX	GNN	SHAP over surrogate graph	Handles high-dimensional graphs	Approximation-dependent	[18]
Counterfactuals	XGNN	GNN	Synthetic graph samples	Direct intervention analysis	Interpretation may be non-unique	[15]
	CF-GNNExplainer	GNN	Graph edit-based scenarios	Minimal-change rationale	Noisy for sparse graphs	[19]
Rule-Based	AMIE+, RDF2Rules	KG Embeddings	Symbolic IF–THEN rules	Interpretable, human-readable	Prone to overfitting rules	[20]
	KGExplainer, SeXAI	KGE + Symbolic Logic	First-order rules	High semantic fidelity	Domain-specific tuning needed	[21]

A survey paper which identifies four key roles of KGs in supporting explainability: (1) KG Construction, structuring and semantically labelling entities; (2) Feature Extraction, identifying meaningful features for AI models; (3) Relationship Extraction, resolving semantic conflicts and enhancing data integration; and (4) KG Reasoning, inferring new facts and generating logical explanations to clarify AI model decisions.

Also, they highlighted how KGs enhance explainability across the AI model lifecycle by supporting three key stages: (1) Pre-modelling, where KGs standardize datasets and extract entities or relationships, such as structuring medical knowledge from electronic health records; (2) In-modelling, where KGs guide model reasoning, as exemplified by using KG embeddings to improve disease classification models; and (3) Post-modelling, where KGs provide transparent explanations through logical reasoning, like generating interpretable, rule-based explanations for drug recommendations.

Lecue argues that KGs serve as a powerful semantic layer to bridge the gap between black-box models and human-understandable reasoning. By encoding context, structure, and causal relationships, KGs enable more meaningful local and global explanations in domains such as machine learning, neural networks, robotics, and natural language processing. The paper emphasizes that combining symbolic reasoning with statistical learning through Knowledge Graphs is a promising direction for building trustable and interpretable AI system [22].

C. Categorizing XAI Methods for Knowledge Graph-based Drug Repurposing

In the context of drug repurposing using KGs, I categorize XAI methods based on two main dimensions: (1) what they explain, Referring to the type of insight provided-such as logical rules, feature attributions, or counterfactuals-and (2) how they generate explanations, meaning the underlying approach used, including rule mining, perturbation techniques, or surrogate models. These dimensions form the basis for the categorization presented in the following section.

1) Graph-Based Feature Attribution

Graph-based XAI methods identify which nodes, edges, or features in a KG contributed most to a model's prediction.

These methods are particularly suited for Graph Neural Networks (GNNs), commonly used in drug repurposing [23].

I. GNNExplainer (Graph Neural Network Explainer)

GNNExplainer is a model-agnostic explanation method designed for Graph Neural Networks (GNNs). It identifies a minimal subgraph and set of node features that are most influential to a specific prediction. For example, in drug repurposing, it can highlight a subgraph showing how Drug A interacts with Gene B, which is involved in Disease C. This provides an interpretable justification for predicting a drug disease link by revealing the critical biological components influencing the model's decision [13].

II. GraphLIME (Graph-based Local Interpretable Model-Agnostic Explanations)

GraphLIME adapts the LIME framework to graph data by learning a local, interpretable linear model around a target node. It perturbs the node's neighbourhood in the graph and uses feature attribution to assess the impact of each node feature on the model's prediction. For example, it can explain why a certain gene is linked to a drug effect by identifying key molecular descriptors or pathway-related features that influence the prediction locally [14].

III. GraphSHAP (Shapley Value-based Explainability for Graphs)

GraphSHAP assigns Shapley values to graph elements-nodes, edges, or features-by evaluating their contribution to the model's prediction across all possible subgraphs. In biomedical applications, it can quantify the impact of entities like proteins, pathways, or molecular substructures on a predicted outcome, such as drug efficacy or disease association. This allows for a precise, theoretically grounded explanation of what made the model reach a specific decision [12].

2) Path-Based Explainability

Path-based XAI methods generate explanations by tracing multi-hop relational paths in KGs, offering semantically rich insights into link prediction models.

I. eXpath (Ontological Closed Path Rules for LP Explanation)

eXpath is a post-hoc explanation framework for embedding-based link prediction. It identifies relation paths between head and tail entities and compresses them into closed path (CP) and property transition (PT) rules. For example, a predicted triple ⟨Drug A, treats, Disease B⟩ can be explained through a path such as Drug A - targets → Gene X - associated_with → Disease B, providing an interpretable ontological rule supporting the prediction [16].

II. Power-Link (GNN-based Path Explanation for KGC)

Power-Link is a scalable path-based explainer for GNN-based knowledge graph completion (KGC) models. It constructs a triplet edge scorer (TES) and employs a graph-powering technique to enhance and extract influential paths. Instead of subgraphs or isolated facts, Power-Link produces paths like City → located_in → Country → part_of → Region to justify the prediction of ⟨City, located_in, Region⟩, aligning with human reasoning and offering efficiency for large-scale KGs.

3) Surrogate explainability methods (Probabilistic Methods)

Create simple, interpretable models that mimic how complex AI systems make decisions. These methods help explain black-box predictions by showing which parts of the data (e.g., nodes or features in a biomedical graph) influenced the output.

I. PGM-Explainer (Probabilistic Graphical Model Explainer)

PGM-Explainer is a surrogate XAI method that uses Bayesian networks to approximate the prediction behaviour of a graph neural network. It learns a probabilistic model over a KG and estimates which nodes and edges are most important for a specific prediction. For example, in a drug repurposing scenario, it may reveal that a drug's effect on a gene and the gene's relation to a disease jointly increase the prediction confidence. This enables explainability with built-in uncertainty estimation, which is particularly valuable in clinical decision-making [17].

II. GraphSVX (Graph-based SHAP Surrogate Model)

GraphSVX combines surrogate modeling with Shapley value theory to explain predictions made by graph-based models. It builds a simplified version of the original graph using random sampling and computes SHAP values to determine the importance of each node or feature. In large biomedical KGs, GraphSVX can identify, for instance, that a combination of gene expression levels and protein interactions strongly influenced a drug-disease prediction, offering interpretability in high-dimensional, noisy data environments [18].

4) Counterfactual & Graph Generation-Based XAI

Counterfactual and graph generation-based XAI methods test "what-if" scenarios by modifying the graph to see how predictions change. These methods help reveal why a model made a decision by exploring how it would behave if key elements were different.

I. XGNN (Explainable Graph Neural Network)

XGNN is a model that generates new graphs to test how a Graph Neural Network (GNN) makes predictions. By systematically adding or removing edges or nodes-like a specific protein interaction, it checks how these changes affect the outcome. For example, a test might involve assessing whether, if the link between Protein X and Disease Y is removed, the drug prediction for Disease Y still holds. This approach facilitates the identification of key features or relationships that influence the model's predictive outcome [15].

II. CF-GNNExplainer (Counterfactual Graph Neural Network Explainer)

CF-GNNExplainer generates counterfactual explanations by modifying nodes or edges in a graph to flip the model's prediction. It asks: What minimal change to the graph would lead the AI to change its decision? For example, if a model predicts that Drug A treats Disease B, CF-GNNExplainer may show that removing a specific gene–disease connection would reverse the prediction. This helps uncover critical dependencies and supports decision auditing in drug discovery [19].

5) Rule-Based Learning for Explainability

Rule-based XAI methods extract logical relationships from KGs, making them highly interpretable

I. AMIE+ (Association Rule Mining in KGs) AMIE+ is a rule-mining algorithm that extracts logical rules from biomedical KGs. For example, it can derive patterns such as:

Drug A treats Disease B → Drug A targets Gene C → Gene C is associated with Disease B. This form of rule-based reasoning helps justify why a particular drug might be predicted for repurposing by making the connections between drugs, genes, and diseases explicit and interpretable [20].

II. RDF2Rules (Frequent Pattern Mining in KGs)

RDF2Rules is a rule-mining approach that identifies hidden relationships in KGs using frequent pattern mining. It uncovers logical rules that link entities based on shared patterns. For example:

IF Drug A has side effect X AND side effect X is also a symptom of Disease B, THEN Drug A may treat Disease B.

This method is particularly effective for side-effect-based drug repurposing, as it generates explicit, human-readable rules that trace meaningful connections between drugs and diseases [24].

III. SeXAI (Semantic Explainable AI)

SeXAI enhances the interpretability of deep learning models by translating their outputs into First-Order Logic (FOL) rules. It leverages ontology-based reasoning to provide structured and biologically meaningful explanations. Designed for applications like drug discovery, SeXAI aligns naturally with the semantic structure of biomedical KGs, making its outputs both explainable and domain-relevant [9].

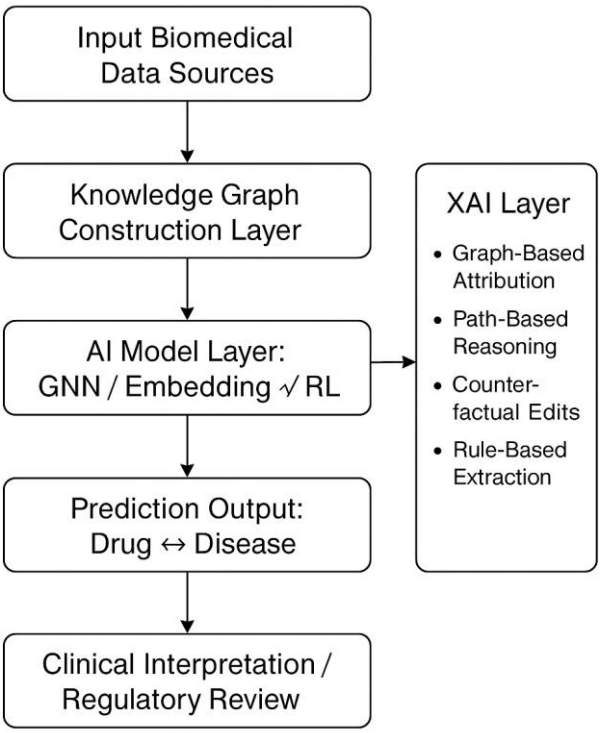

Fig. 1: Framework for Integrating XAI into KG-Based Drug Repurposing

6) XAI Approaches

In biomedical AI, especially in complex tasks like drug repurposing using KGs, combining multiple XAI techniques can significantly improve both interpretability and transparency. Rather than relying on a single method, integrating path-based explanations, rule-based reasoning, and feature attribution allows for more comprehensive and trustworthy insights into model predictions.

A powerful hybrid XAI approach merges path-based explainability with rule-based logic, ensuring that AI-driven predictions are not only accurate but also understandable. By leveraging Knowledge Graph Embeddings (KGE), logical rule extraction, and multi-hop relational reasoning, this combined strategy makes the decision-making process in drug repurposing more transparent, explainable, and actionable for both researchers and clinicians.

For instance, the KGML-xDTD model offers a path-based explanation framework by integrating reinforcement learning, KGs, and adversarial training to identify interpretable paths between drugs and diseases. This approach enhances user trust by addressing black-box concerns and providing human-understandable rationales for predictions in biomedical research [25].

Another example is the application of rule-based XAI in explainable COVID-19 drug prediction, where logical rules are extracted from KGs to clarify how AI links drugs to diseases. This method combines rule mining with path-based reasoning, tracing connections through intermediary biological entities like genes and proteins. It results in biologically meaningful and clinically relevant explanations for drug–disease associations [26].

Additionally, KGExplainer is a rule-based XAI method specifically designed for Knowledge Graph Embeddings (KGEs). It improves interpretability by extracting symbolic rules from high-dimensional latent representations. KGExplainer follows a structured five-step process to detect statistical patterns in subgraph neighbourhoods and produces rule-based, instance-based, and analogy-based explanations. It incorporates surrogate models such as MDI, K-Lasso, and HSIC-Lasso to select the most informative rules. Importantly, it does this without retraining the original model, making it a scalable and faithful solution for large biomedical KGs [26].

ExamplE (Explanation Graph) is another embedding-based XAI approach that explains predictions by identifying influential training examples in the latent space of KGE models. It reconstructs semantic explanation subgraphs that highlight how similar examples contribute to a prediction, offering interpretable evidence for link prediction tasks. This method is particularly useful for illustrating semantic relevance in biomedical reasoning without altering the original embedding model [27].

Complementing these approaches, TxGNN is a graph-based foundation model for clinician-centred drug repurposing that excels in zero-shot scenarios involving rare or poorly understood diseases. It integrates metric learning with GNNs and provides multi-hop explanations via a dedicated explainer module, directly leveraging medical KGs to enhance clinical trust and decision-making [28].

Similarly, rd-explainer focuses on rare disease drug repurposing by combining disease-specific KGs and GNNs with the GNNExplainer framework. It delivers high-quality predictions supported by semantic subgraph explanations,

ensuring interpretability and alignment with biomedical knowledge [29].

To illustrate the practical value of the proposed methods, this study examines a real-world case involving Metformin and Alzheimer's

TABLE II. MAPPING XAI TECHNIQUES TO STAKEHOLDER NEEDS

Stakeholder	Explanation Needs	Appropriate XAI Techniques	Justification
Clinician	Causal or semantic interpretability	Rule-based (AMIE+), Path-based (eXpath)	Aligns with clinical logic and treatment pathways
AI Researcher	Feature contribution, debugging	GraphSHAP, GNNExplainer	Quantifies internal model behavior
Regulator	Transparent, auditable, consistent outputs	Surrogate Models (PGM-Explainer), KGExplainer	Supports compliance and traceability

III. CASE STUDY - INTERPRETING DRUG–DISEASE PREDICTIONS WITH XAI

A. Objective

While this paper presents a comprehensive taxonomy of XAI methods applicable to drug repurposing, it is essential to substantiate their value through a concrete case study. This section presents a focused experimental demonstration that illustrates how an XAI method - GNNExplainer can elucidate the reasoning behind a machine learning model's prediction of a novel therapeutic association within a biomedical KG. Specifically, this study explores the predictive relationship between the drug Metformin and Alzheimer's Disease using the Hetionet KG [19].

B. Methodology

1) KG and Data Preparation

Hetionet [30] was selected for this case study due to its rich, heterogeneous biomedical structure, integrating data on drugs, diseases, genes, pathways, and compounds from 29 public databases. The graph includes over 47,000 nodes and 2 million edges. For the purpose of link prediction, a subset of Hetionet containing drug–gene–disease paths with known therapeutic relevance was extracted.

2) Model and Task

A basic Graph Convolutional Network (GCN) was trained to predict the existence of a therapeutic link between drug–disease pairs. The task was framed as a binary classification problem, where positive instances were known therapeutic links and negative instances were sampled from unlinked drug - disease pairs [31, 32].

3) XAI Method: GNNExplainer

GNNExplainer was employed post-hoc on a model-predicted positive link:

Metformin → Alzheimer's Disease

This prediction, supported by existing biomedical literature, serves as a representative use case for evaluating XAI methodologies in drug repurposing. GNNExplainer was configured to extract the minimal subgraph and set of node features that contributed most significantly to the model's prediction.

C. Results

The explainer returned a biologically meaningful path:

Fig. 2: Framework for Integrating XAI into KG-Based Drug Repurposing

The explanation subgraph generated by GNNExplainer is consistent with published pharmacological studies, thereby supporting the plausibility of the model's prediction, indicating that Metformin modulates SIRT1, which has neuroprotective effects and is involved in amyloid-beta and tau regulation [13].

D. Discussion

This case study validates the argument that XAI, particularly graph-based attribution techniques, can render AI-driven drug repurposing interpretable and biologically plausible. The use of GNNExplainer allowed us to generate transparent, subgraph-based rationales that link molecular interactions to high-level disease outcomes.

Importantly, this approach:

- Enhances trust and regulatory acceptability of black-box AI models.
- Aids researchers and clinicians in hypothesis validation.
- Reveals semantic paths that can be targeted for further wet-lab experimentation.

While this demonstration focuses on a single link, the same methodology is extensible to larger-scale inference tasks. Limitations include GNNExplainer's sensitivity to graph topology noise and computational cost, which may be mitigated in future work through path-based or rule-based XAI models.

IV. EVALUATING XAI IN KG-BASED DRUG REPURPOSING

A. Current Metrics in Practice

A number of quantitative and qualitative metrics are currently used to evaluate XAI methods across domains. Table III summarizes key metrics relevant to graph-based and biomedical contexts.

B. Gaps in Current Metrics

While many of these metrics are informative, they were developed for tabular or image data and may not fully address the semantically rich, multi-relational structure of biomedical KGs. For example, sparsity in a graph context may not always imply better interpretability if meaningful multi-hop connections are suppressed.

C. *Proposed Domain-Aware Composite Evaluation Framework*

A multi-metric framework combining intrinsic and extrinsic evaluation for drug repurposing is proposed:

- Intrinsic Metrics:
  - Fidelity (e.g., explanation consistency with prediction probability drop)
  - Sparsity (e.g., size of explanation subgraph)
  - Stability (e.g., explanation Jaccard index over perturbed graph instances)
- Extrinsic Metrics:
  - Biomedical Plausibility Score: Overlap with known drug–gene–disease triplets from curated datasets (e.g., DrugBank, CTD)
  - Expert Validation Agreement: Concordance with domain expert rating on relevance/usefulness (Likert-scale-based)
  - Explanation Utility in Downstream Task: Improvement in expert-driven hypothesis formulation or ranking

This framework can be implemented using combinations of model outputs, ontology-based similarity tools (e.g., UMLS, GO), and domain expert feedback, thus creating a pipeline that is both quantifiable and clinically meaningful.

V. Role-Aware Explainability - Adapting XAI to Stakeholder Needs

This paper proposes a mapping of stakeholder roles to explanation types, goals, and preferred presentation formats, as illustrated in Table IV.

TABLE III. Evaluation Metrics for Graph-Based XAI in Drug Repurposing

Metric	Definition	Relevance to KG Drug Repurposing
Fidelity	Degree to which the explanation matches the model's actual behavior	Ensures biological justifications reflect model logic
Sparsity	Number of features, nodes, or edges in the explanation	Aids interpretability by reducing cognitive burden
Stability	Consistency of explanations for similar inputs or perturbations	Avoids contradictory rationales for similar drug-disease pairs
Comprehensibility	Ease with which a domain expert understands the explanation	Crucial in translating outputs to clinical insights
Biological Plausibility	Alignment with known biomedical pathways or literature	Supports external validation against scientific evidence
Localization Accuracy	Precision in identifying the causally responsible graph substructure	Ensures explanation components are not arbitrary or misleading

TABLE IV. Mapping XAI Outputs to Stakeholder Needs in Drug Repurposing

Stakeholder Role	Primary Objective	Explanation Needs	Preferred XAI Techniques
Clinician	Ensure patient safety and treatment rationale	Causal, concise, biologically plausible pathways	Path-based (e.g., eXpath), Rule-based (AMIE+)
Biomedical Researcher	Hypothesis generation, biomarker discovery	Multi-hop interactions, gene/protein roles	GNNExplainer, KGExplainer, GraphSHAP
ML Engineer	Debug models, optimize interpretability	Feature attribution, fidelity, local-global behavior	GraphLIME, GraphSVX, Counterfactuals
Regulator	Validate model traceability and fairness	Transparent, auditable, consistent reasoning	Surrogate models (PGM-Explainer), Rule-based

In practical deployments, a single XAI module should be capable of generating multi-modal outputs, adapting to the audience's profile. This may involve:

- Interface personalization (user-selectable views: logical, visual, statistical).
- Role-based rendering (e.g., "clinical summary" vs. "developer debug mode").
- Integrated explanation dashboards (combining ontology links, saliency, and causal chains).

Emerging research in Human-Centered XAI (HCXAI) and Explanation Personalization supports such adaptive models, emphasizing that interpretability is not universal - it is user-contingent and must be operationalized accordingly [10].

VI. Discussion and Future work

Future research should prioritize the seamless integration of KG construction and XAI to advance the effectiveness and trustworthiness of drug repurposing systems. For KG construction, efforts must focus on developing scalable, distributed architectures capable of handling the rapid expansion of biomedical data while ensuring high data quality, completeness, and consistency. Semi-automated update mechanisms-guided by human-in-the-loop validation-will be crucial for maintaining relevance in dynamic domains such as pharmacogenomics and disease progression. In parallel, XAI approaches must evolve to provide interpretable, secure, and computationally efficient explanations tailored to end-users ranging from clinical researchers to regulatory experts. This includes establishing standardized evaluation frameworks for measuring explanation quality, incorporating privacy-preserving techniques to mitigate data leakage risks, and building adaptive explanation systems that adjust outputs based on user expertise and context. Furthermore, addressing the high computational cost of current explanation techniques will require novel approximation strategies that maintain fidelity without sacrificing performance. By co-developing these capabilities, future systems may enhance the discovery of therapeutic insights by providing interpretable and semantically grounded predictions, pending further empirical validation. But also provide transparent, evidence-driven

justifications necessary for real-world adoption in healthcare and biomedical research.

To facilitate the end-to-end application of XAI in KG-based drug repurposing, a system architecture is presented, comprising four primary layers: data ingestion and integration, KG construction and management, AI model inference, and explanation delivery [33].

TABLE V. SYSTEM MODULES AND RESPONSIBILITY MAPPIN

Module	Description
Data Ingestion Layer	Gathers multi-source biomedical data: DrugBank, ChEMBL, CTD, PubMed, etc. Supports structured, semi-structured, and unstructured formats.
Knowledge Graph Builder	Performs entity extraction, ontology alignment, and KG schema generation. Uses NLP, NER, and relation extraction models.
Graph Storage Engine	Stores KG in graph databases (e.g., Neo4j, RDF triple store). Enables queryability and updates.
4. AI/ML Prediction Engine	Applies GNNs or KGE methods to predict drug–disease links. Supports node classification or link prediction tasks.
XAI Module	Generates human-understandable explanations. Incorporates methods such as GNNExplainer, eXpath, RDF2Rules.
Output Interface	Visual and textual rendering of predictions and explanations tailored to stakeholders (clinicians, researchers, regulators).

Challenges of XAI for Drug Repurposing

Several pressing challenges continue to hinder the effective implementation of XAI in biomedical and KG applications.

1. Lack of standard metrics remains a major issue, as there is no universally accepted method to evaluate explainability, leading to inconsistent comparisons across models and domains. Without a shared framework for measuring qualities like fidelity, interpretability, or usefulness, both researchers and practitioners struggle to benchmark or trust explanation outputs.
2. Security and privacy concerns also emerge when implementing XAI techniques. While transparency is a strength of XAI, it can inadvertently expose sensitive data or increase the system's vulnerability to adversarial attacks, particularly in high-stakes domains such as healthcare.
3. Audience-specific explanation needs further complicate the deployment of XAI systems. Different stakeholders-clinicians, researchers, or regulators-require distinct types of explanations. Current XAI frameworks often fail to tailor outputs to user roles, leading to reduced utility or misinterpretation.
4. The computational cost of generating explanations presents a scalability challenge. Many state-of-the-art XAI methods require multiple iterations, heavy sampling, or surrogate modeling, which can be impractical for large-scale, real-time applications. Addressing these challenges is essential to building XAI systems that are not only transparent but also secure, efficient, and aligned with the needs of diverse end users.

VII. CONCLUSION

The integration of Explainable Artificial Intelligence (XAI) with Knowledge Graphs (KGs) represents a transformative advancement in the domain of drug repurposing. This paper has presented a comprehensive taxonomy of XAI methods applicable to KG-based biomedical inference, alongside a stakeholder-centred interpretability framework and a domain-specific evaluation strategy. Through detailed case analysis and technical comparisons, it has been demonstrated how graph-based, rule-based, and hybrid XAI techniques can bridge the gap between black-box AI models and clinically meaningful insights.

Methods such as GNNExplainer, RDF2Rules, and KGExplainer contribute to enhancing model transparency, a prerequisite for regulatory interpretability and trustworthiness in clinical applications, scientific reproducibility, and hypothesis generation in biomedical research. Moreover, emphasis is placed on the need for role-aware, context-sensitive explanation systems that adapt outputs based on user profiles - clinicians, researchers, engineers, or regulators - thereby enhancing practical utility across disciplines.

Looking ahead, the development of scalable, privacy-preserving, and computationally efficient XAI frameworks remains a pressing challenge. Equally important is the standardization of explanation quality metrics that are both interpretable and biologically grounded. By advancing these directions, future systems will not only accelerate the discovery of novel therapeutic applications but also support safe, ethical, and transparent deployment of AI in healthcare environments.

[1] J. Chen *et al.*, "Knowledge graphs for the life sciences: recent developments, challenges and opportunities. arXiv 5," *arXiv preprint arXiv:2309.17255,* 2023.

[2] P. Perdomo-Quinteiro and A. Belmonte-Hernández, "Knowledge Graphs for drug repurposing: a review of databases and methods," *Briefings in Bioinformatics,* vol. 25, no. 6, p. bbae461, 2024.

[3] X. Pan *et al.*, "Deep learning for drug repurposing: Methods, databases, and applications," *Wiley interdisciplinary reviews: Computational molecular science,* vol. 12, no. 4, p. e1597, 2022.

[4] D. Sun, W. Gao, H. Hu, and S. Zhou, "Why 90% of clinical drug development fails and how to improve it?," *Acta Pharmaceutica Sinica B,* vol. 12, no. 7, pp. 3049-3062, 2022.

[5] S. Li, K. W. Wong, D. Zhu, and C. C. Fung, "Drug-CoV: a drug-origin knowledge graph discovering drug repurposing targeting COVID-19," *Knowledge and Information Systems,* vol. 65, no. 12, pp. 5289-5308, 2023.

[6] F. Yang *et al.*, "Machine learning applications in drug repurposing," *Interdisciplinary Sciences: Computational Life Sciences,* vol. 14, no. 1, pp. 15-21, 2022.

[7] C. Königs, M. Friedrichs, and T. Dietrich, "The heterogeneous pharmacological medical biochemical network PharMeBINet," *Scientific Data,* vol. 9, no. 1, p. 393, 2022.

[8] P. Chandak, K. Huang, and M. Zitnik, "Building a knowledge graph to enable precision medicine," *Scientific Data,* vol. 10, no. 1, p. 67, 2023.

[9] S. Schramm, C. Wehner, and U. Schmid, "Comprehensible artificial intelligence on

knowledge graphs: A survey," *Journal of Web Semantics,* vol. 79, p. 100806, 2023.

[10] M. Mersha, K. Lam, J. Wood, A. AlShami, and J. Kalita, "Explainable artificial intelligence: A survey of needs, techniques, applications, and future direction," *Neurocomputing,* p. 128111, 2024.

[11] P. Gohel, P. Singh, and M. Mohanty, "Explainable AI: current status and future directions," *arXiv preprint arXiv:2107.07045,* 2021.

[12] A. Perotti, P. Bajardi, F. Bonchi, and A. Panisson, "Explaining identity-aware graph classifiers through the language of motifs," in *2023 International joint conference on neural networks (IJCNN)*, 2023: IEEE, pp. 1-8.

[13] Z. Ying, D. Bourgeois, J. You, M. Zitnik, and J. Leskovec, "Gnnexplainer: Generating explanations for graph neural networks," *Advances in neural information processing systems,* vol. 32, 2019.

[14] Q. Huang, M. Yamada, Y. Tian, D. Singh, and Y. Chang, "Graphlime: Local interpretable model explanations for graph neural networks," *IEEE Transactions on Knowledge and Data Engineering,* vol. 35, no. 7, pp. 6968-6972, 2022.

[15] H. Yuan, J. Tang, X. Hu, and S. Ji, "Xgnn: Towards model-level explanations of graph neural networks," in *Proceedings of the 26th ACM SIGKDD international conference on knowledge discovery & data mining*, 2020, pp. 430-438.

[16] Y. Sun, L. Shi, and Y. Tong, "eXpath: Explaining Knowledge Graph Link Prediction with Ontological Closed Path Rules," *arXiv preprint arXiv:2412.04846,* 2024.

[17] M. Vu and M. T. Thai, "Pgm-explainer: Probabilistic graphical model explanations for graph neural networks," *Advances in neural information processing systems,* vol. 33, pp. 12225-12235, 2020.

[18] A. Duval and F. D. Malliaros, "Graphsvx: Shapley value explanations for graph neural networks," in *Machine Learning and Knowledge Discovery in Databases. Research Track: European Conference, ECML PKDD 2021, Bilbao, Spain, September 13–17, 2021, Proceedings, Part II 21*, 2021: Springer, pp. 302-318.

[19] A. Lucic, M. A. Ter Hoeve, G. Tolomei, M. De Rijke, and F. Silvestri, "Cf-gnnexplainer: Counterfactual explanations for graph neural networks," in *International Conference on Artificial Intelligence and Statistics*, 2022: PMLR, pp. 4499-4511.

[20] L. Galárraga, C. Teflioudi, K. Hose, and F. M. Suchanek, "Fast rule mining in ontological knowledge bases with AMIE $$+ $$," *The VLDB Journal,* vol. 24, no. 6, pp. 707-730, 2015.

[21] A. Gan *et al.*, "Retrieval Augmented Generation Evaluation in the Era of Large Language Models: A Comprehensive Survey," *arXiv preprint arXiv:2504.14891,* 2025.

[22] F. Lecue, "On the role of knowledge graphs in explainable AI," *Semantic Web,* vol. 11, no. 1, pp. 41-51, 2020.

[23] Z. Wu, S. Pan, F. Chen, G. Long, C. Zhang, and P. S. Yu, "A comprehensive survey on graph neural networks," *IEEE transactions on neural networks and learning systems,* vol. 32, no. 1, pp. 4-24, 2020.

[24] Z. Wang and J. Li, "RDF2Rules: Learning rules from RDF knowledge bases by mining frequent predicate cycles," *arXiv preprint arXiv:1512.07734,* 2015.

[25] C. Ma, Z. Zhou, H. Liu, and D. Koslicki, "KGML-xDTD: a knowledge graph–based machine learning framework for drug treatment prediction and mechanism description," *GigaScience,* vol. 12, p. giad057, 2023.

[26] M. K. Islam, D. Amaya-Ramirez, B. Maigret, M.-D. Devignes, S. Aridhi, and M. Smaïl-Tabbone, "Molecular-evaluated and explainable drug repurposing for COVID-19 using ensemble knowledge graph embedding," *Scientific Reports,* vol. 13, no. 1, p. 3643, 2023.

[27] A. Janik and L. Costabello, "Explaining Link Predictions in Knowledge Graph Embedding Models with Influential Examples," *arXiv preprint arXiv:2212.02651,* 2022.

[28] K. Huang *et al.*, "A foundation model for clinician-centered drug repurposing," *Nature Medicine,* vol. 30, no. 12, pp. 3601-3613, 2024.

[29] P. Perdomo-Quinteiro, K. Wolstencroft, M. Roos, and N. Queralt-Rosinach, "Knowledge graphs and explainable ai for drug repurposing on rare diseases," *bioRxiv,* p. 2024.10. 17.618804, 2024.

[30] D. S. Himmelstein *et al.*, "Systematic integration of biomedical knowledge prioritizes drugs for repurposing," *Elife,* vol. 6, p. e26726, 2017.

[31] R. Yue and A. Dutta, "Repurposing Drugs for Infectious Diseases by Graph Convolutional Network with Sensitivity-Based Graph Reduction," *Interdisciplinary Sciences: Computational Life Sciences,* vol. 17, no. 1, pp. 185-199, 2025.

[32] M. Schlichtkrull, T. N. Kipf, P. Bloem, R. Van Den Berg, I. Titov, and M. Welling, "Modeling relational data with graph convolutional networks," in *The semantic web: 15th international conference, ESWC 2018, Heraklion, Crete, Greece, June 3–7, 2018, proceedings 15*, 2018: Springer, pp. 593-607.

[33] K. PETROVIČOVÁ, "Knowledge Graphs and Explainable Predictive Models for Drug Repurposing," MASARYK University, 2023.