Extending the Nested Model for User-Centric XAI: A Design Study on GNN-based Drug Repurposing

Qianwen Wang, Kexin Huang, Payal Chandak, Marinka Zitnik, Nils Gehlenborg

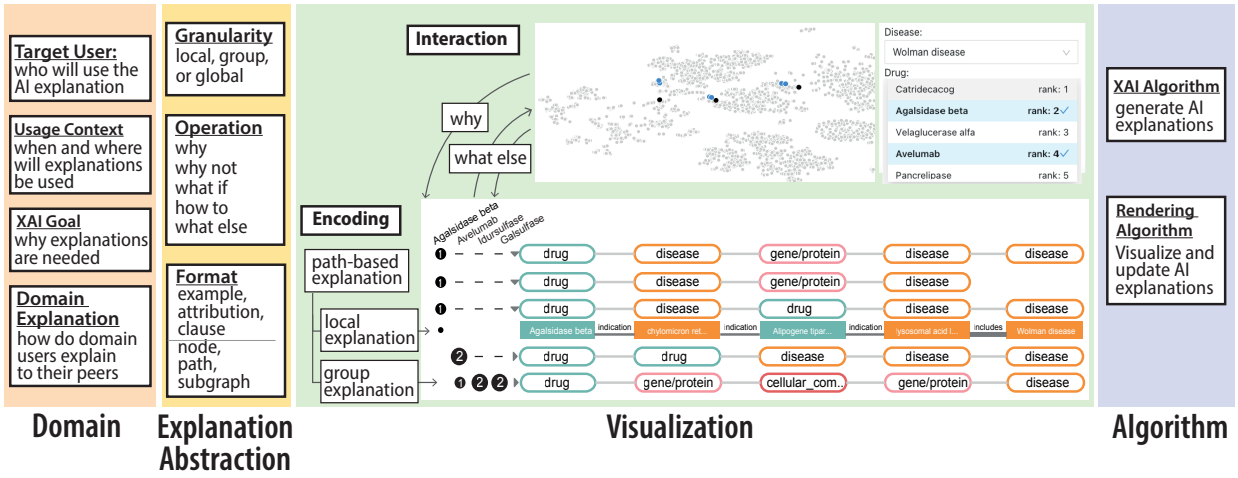

Fig. 1. We design and develop DrugExplorer for domain users to understand and assess graph neural network-based drug repurposing. The design process follows the nested model of visualization design and extends it by adding user-centric XAI design considerations. As in the nested block and guideline model (NBGM) [39], the four nested layers are drawn separately for visual simplicity.

Abstract— Whether AI explanations can help users achieve specific tasks efficiently (*i.e.*, usable explanations) is significantly influenced by their visual presentation. While many techniques exist to generate explanations, it remains unclear how to select and visually present AI explanations based on the characteristics of domain users. This paper aims to understand this question through a multidisciplinary design study for a specific problem: explaining graph neural network (GNN) predictions to domain experts in drug repurposing, *i.e.*, reuse of existing drugs for new diseases. Building on the nested design model of visualization, we incorporate XAI design considerations from a literature review and from our collaborators' feedback into the design process. Specifically, we discuss XAI-related design considerations for usable visual explanations at each design layer: target user, usage context, domain explanation, and XAI goal at the domain layer; format, granularity, and operation of explanations at the abstraction layer; encodings and interactions at the visualization layer; and XAI and rendering algorithm at the algorithm layer. We present how the extended nested model motivates and informs the design of DrugExplorer, an XAI tool for drug repurposing. Based on our domain characterization, DrugExplorer provides path-based explanations and presents them both as individual paths and meta-paths for two key XAI operations, *why* and *what else*. DrugExplorer offers a novel visualization design called *MetaMatrix* with a set of interactions to help domain users organize and compare explanation paths at different levels of granularity to generate domain meaningful insights. We demonstrate the effectiveness of the selected visual presentation and DrugExplorer as a whole via a usage scenario, a user study, and expert interviews. From these evaluations, we derive insightful observations and reflections that can inform the design of XAI visualizations for other scientific applications.

Index Terms—Visual Explanation, XAI, Visualization Design Model, Drug Repurposing

1 Introduction

Recent years witnessed a rapid expansion of Artificial Intelligence (AI) techniques in various domains and a growing need for eXplainable Artificial Intelligence (XAI). While a variety of algorithms have been proposed to generate explanations, there is no guarantee that these explanations are always usable in the applied domain, *i.e.*, whether domain users can use AI explanations to complete desired tasks efficiently. Even though some studies demonstrate the positive effects of AI explanations [38], others report that AI explanations fail to generate actionable insights and even manipulate user trust [1, 75]. Recently, AI researchers started to recognize usability as an indispensable requirement for AI explanations [59, 87].

Usable AI explanations not only require accurate, stable, and faithful algorithms, but also need well-designed user interfaces that bridge the capabilities of algorithms to the needs of users in application domains [15, 36, 68, 69]. Researchers have advocated for user-centered XAI, within which usable explanations are extensively discussed [15, 36, 40, 68, 69]. These studies provide valuable frameworks and guidelines for designing explanation interfaces, either by borrowing lessons from social science and psychology [40, 68] or conducting empirical studies with real users [15, 36]. However, these studies mainly discuss the design of general user interfaces without a specific investigation about interactive visualizations, which is a crucial component in explanation interfaces.

Given the importance of visualization in AI, a growing number of AI visualization tools have been proposed. Most existing AI visu-

- *Qianwen Wang, Marinka Zitnik, Nils Gehlenborg are with Harvard University. E-mail: {qianwen_wang, marinka, nils}@hms.harvard.com.*
- *Kexin Huang is with Stanford University. E-mail: kexinh@stanford.edu.*
- *Payal Chandak is with Harvard-MIT Health Sciences and Technology. E-mail: payal_chandak@hst.harvard.edu.*

Manuscript received xx xxx. 201x; accepted xx xxx. 201x. Date of Publication xx xxx. 201x; date of current version xx xxx. 201x. For information on obtaining reprints of this article, please send e-mail to: reprints@ieee.org. Digital Object Identifier: xx.xxxx/TVCG.201x.xxxxxxx

alization tools are developed for AI developers and AI practitioners rather than domain users who have limited expertise in AI [82]. Studies that target domain users [14, 72] often concentrate on proposing novel visualization designs and coordinated views to make sense of complicated data. A specific explanation is usually selected before the design study based on its popularity in the ML community without considering how the domain characteristics and user needs may influence the selection and visualization of explanations. However, many user studies demonstrated that the visual presentation of explanations could significantly influence humans in using AI, ranging from confidence level to performance accuracy [4, 8, 18]. While the visualization field has accumulated extensive experience in developing visualization tools for domain users and summarized many insightful visualization models [11, 39, 43, 57, 65], the visualization designs are mainly driven by domain problems. It remains unclear about how to effectively investigate and fulfill users' needs for usable explanations through the visualization design process.

This paper presents a design study where we investigated how to select and visualize AI explanations for domain users. We focus on one particular scientific application of AI, *i.e.*, graph neural networks (GNN) in drug repurposing, which enables us to closely work with both domain and AI experts, iterate designs based on user feedback, and conduct evaluations on real datasets and tasks. Our design study follows the nested model and the nested block and guideline model (NBGM) [39, 43] [1] since they provide explicit mechanisms to capture and justify design decision rationales. We incorporate a diverse set of user-centric XAI design considerations into different layers of the design processes, based on our literature review and collaborators' feedback in our iterative design study, as shown in Fig. 1. **This design process decouples the explanation abstraction from the XAI algorithm, aiming to provide explanation visualizations and interactions that better reflect the domain characterization.** Based on the domain characterization (target user, usage context, XAI goal, domain explanation), DrugExplorer provides path-based explanations and presents them both at both instance level and group level for two key XAI operations, *why* and *what else*. We also propose a novel visualization design *MetaMatrix* to help domain users organize and compare explanation paths at different levels of granularity to generate domain-meaningful insights for their XAI goals.

This paper makes three main contributions:

- We design and develop an interactive visualization tool, DrugExplorer, to assist domain users in GNN-based drug repurposing.
- We present the design process of DrugExplorer, which applies the nested model to the selection and visualization of AI explanations by incorporating a diverse set of user-centric XAI considerations.
- We evaluate DrugExplorer and share observations and insights that are generalizable and valuable for the development of future domain-specific XAI visualization tools.

2 Study Context: Graph Neural Networks in Drug Repurposing

Drug repurposing is an effective strategy to identify new therapeutic uses of existing drugs. Compared to developing a new drug from scratch, which typically takes 13-15 years and 2-3 billion dollars on average, repurposed drugs can potentially get to market in half the time and at one-quarter of the cost [45]. However, despite considerable advances, current examples of successful drug repurposing mainly came about through serendipity.

Recently, GNNs have emerged as a promising approach in computational drug repurposing. However, predicted candidate drugs require rigorous and systematic validation, including in vitro experiments, in vivo experiments, and clinical trials. Given limited resources, a critical task for domain experts is to decide which candidate drugs to investigate further and which ones to leave out.

[1] Unless specified otherwise, the nested model refers to both its original version and its extension: the nested block and guideline model (NBGM).

3 Related Work

3.1 User-centric XAI

User-centric XAI investigates how humans interpret, interact with, and use XAI. Here, we review user-centric XAI studies that inform the design of explanation interfaces for non-AI-expert end users.

To guide the design of explanation interfaces, some researchers empirically study users' behavior and needs when using XAI. For example, Chen *et al.* [15] found that interactive explanations were more effective than static ones at improving user comprehension, but with the cost of longer decision time. Feng and Boyd-Craber [18] observed that users used different game playing strategies with highlight-, guess-, and evidence-based explanations. On the other hand, by conducting case studies and expert interviews, Zytek *et al.* [87] summarized a list of usability challenges of AI in high-stakes decision making. Liao *et al.* [36] present user needs for explainability as a set of prototypical questions.

Another parallel research aims to summarize guidelines and form frameworks by reviewing literature in related files such XAI, human-AI interaction, psychology, and social science [6, 12, 34, 42, 58, 68]. For example, Chari *et al.* [12] proposed Explanation Ontology, which can help designers identify the components that an XAI system should and can provide to its end users. Mohseni *et al.* [42] presented a framework that categorizes the design goals of XAI and provides guidelines to evaluate these goals at each stage of the design process. Most relevant to our study is the conceptual framework contributed by Wang *et al.* [68]. This framework maps algorithm-generated explanations to human decision-making theories, aiming to mitigate bias decision errors by helping users select appropriate explanation types and representations for domain-specific XAI applications. Despite the valuable guidelines provided, the proposed framework failed to include how the characteristics of the domain problem can influence the selection of AI explanations. Moreover, the framework only covered simple visualizations (*e.g.*, bar charts, heat map) and provides limited guidance for the typical multi-step visualization design process.

Focusing on visualization design for usable explanations, this paper adapts the nested model [39, 43] to design usable visual explanation interfaces. We discuss the threats and validation methods for usable visual explanations at each level of the design framework.

3.2 Visualizations for XAI

Interactive visualizations have been widely used as a medium for explanation [14, 23, 82], since they excel at communication and summarization of complex information.

Most existing AI visualization tools are developed for AI developers and AI practitioners [23, 82]. These tools succeed on a range of tasks, including data augmentation and cleaning [13, 78], model debugging [10], and model comparison and selection [71–73]. However, domain users have different expertise and analysis goals than AI experts. As a result, these tools can generally not be directly applied for domain users.

Some recent studies take into account the needs of domain users for the development of XAI visualization tools [14, 33, 41]. Theses studies contribute novel visualization designs and coordinated views to help domain users make sense of complicated data and generate domain-meaningful insights. However, these tools usually employed one particular explanation technique and representation selected based on either popularity or state-of-the-art. As a result, they don't consider the selection of explanations in the design process. However, these tools usually left the selection of XAI outside the design process, choosing an explanation technique and representation based on its popularity in the ML community. Furthermore, existing visualization models [11, 39, 43, 57, 65] propose no explicit design guidelines for AI explanation selection and representation.

This study incorporates the selection of explanation techniques and representations into the design of an XAI visualization tool. We extend the nested design model, with a particular focus on how the characteristics of the domain problem shape the selection of explanations and the design of visualizations.

3.3 AI and Visualization in Drug Repurposing

Recent advances in AI have presented impressive capabilities to repurpose drugs at unprecedented speed, scale, and accuracy. AI-assisted drug repurposing attracts increasing research interest, especially for treating emerging and challenging diseases, such as COVID-19 [21]. A widely-used AI model for drug repurposing are GNNs. Many research efforts have been undertaken in GNN-based drug repurposing, including construction of knowledge graphs that comprehensively summarize the existing biomedical knowledge [25, 84], and development of GNN models that can effectively learn from large knowledge graphs [21, 60].

Even though current GNN-based drug repurposing approaches show promising performance, they usually provide limited explanations, which are important to validate new findings and extend human understanding of how drugs act in different diseases [27]. This gap partly comes from the complicated nature of drug discovery, as well as the challenges of conducting multidisciplinary research across the fields of visualization, biomedicine, AI, and human-computer interaction. A vast array of XAI techniques have been proposed to generate explanations for GNN predictions [2,55,79]. Meanwhile, many visual analytics tools have been proposed to present complex biological pathways and assist domain users in drug discovery [32, 35, 47–49, 54]. However, how to combine these XAI and visualization techniques to facilitate human-AI collaboration in drug repurposing is still an open question.

This study builds upon prior studies in GNN-based drug repurposing and GNN explainability. While the visualization design is largely inspired by previous studies on visual analytics of graphs and biological pathways, our focus is on defining the designing process for visualizing AI explanations for domain experts.

4 Incorporating XAI Design Considerations into the Nested Model of Visualization Design

This section introduces the motivation and the methodology for incorporating XAI considerations into the visualization design process.

4.1 Overview

Our design study included two stages and was conducted by a multidisciplinary team with diverse backgrounds in visualization, XAI, and biomedicine. In the first stage, we investigated the needs and challenges in explaining AI-based drug repurposing to domain experts with a prototype tool on a specific disease, SARS-CoV-2 (Figure S3, Supplementary Material). This prototype was driven by an initial set of requirements and an explanation method (*i.e.*, GNNExplainer [79]) provided by the XAI researchers, who are also co-authors of this paper. In the second stage, we designed and developed DrugExplorer for general drug repurposing based on the feedback in stage one. The team met on a regular basis to discuss the visualization results of GNN explanations and predictions and iterate the design based on expert feedback.

The design study reveals the challenge of directly applying existing design study models. In particular, current visualization models are usually driven by domain problems and provide little guidance about 1) how to investigate experts' needs for AI explanations and 2) how these needs influence visualization design decisions. To incorporate XAI considerations in our visualization design, we choose the nested model among existing visualization models to highlight design decisions rather than the design process (*e.g.*, [57]) or architecture (*e.g.*, [11, 65]). The nested model provides a clear structure to describe and justify design decisions (*i.e.*, design decisions are categorized into four nested layers and connected via design guidelines). Therefore, it serves as a useful backbone structure to incorporate XAI design considerations into our design decisions. To guide the design process with representative XAI considerations, we first extracted all the XAI-related design considerations from nine XAI design frameworks [12, 17, 19, 24, 28, 34, 36, 42, 68]. We then merged similar design considerations and removed design considerations that are not related to domain users (*e.g.*, design considerations targeted at AI novices). We carefully fit these design considerations into the four layers of the nested model. We tested and modified these considerations throughout the design process based on users' feedback about DrugExplorer. Table 1 summarizes how different XAI considerations are incorporated into the design process and extend the nested model.

Table 1. User-centric XAI considerations in the visualization design.

Block	Notes	Ref
Domain		
Target User	users' research field, AI expertise, and role in using AI systems	[12, 19, 24, 34, 36, 42]
Usage Context	when and where will the AI explanations be used (*e.g.*, time sensitivity)	
XAI Goal	domain-related problems that the users aim to solve using AI explanations	
Domain Explanation	how a human expert would reason about a phenomenon in the applied domain	user mental model in [17, 24, 42]
Abstraction		
Format	**attribution:** explain using feature attributes (*e.g.*, salience map, feature importance scores) **example:** explain using similar or contrastive examples **clause:** explain using rules or decision trees	[28, 68] mainly for Euclidean data such as images and tables
	node: important neighbor nodes to the prediction targets **path:** important message passing for the prediction targets **subgraph:** important subgraphs around the prediction targets	Our survey on GNN explanations based on [80]
Granularity	**local:** explain an individual prediction **global:** explain a prediction process of a model	[34, 36, 42]
	group: explain a group of similar predictions	collaborators' feedback
Operation	***why:*** reason about why a certain prediction is made ***why not:*** reason about why a certain prediction is not made ***what if:*** understand how a specific modification will influence the prediction ***how to:*** investigate the adjustment needed to generate a different prediction ***what else:*** query similar instances that generate similar predictions	[36, 42, 68]
Visualization		
Encoding	how to present the explanation **format** at selected levels of **granularity**	partially covered in [28]
Interaction	how to support required **operations** at the selected explanation **format**	

4.2 Domain

In the domain characterization layer, a visualization designer identifies the domain problems and needs related to the design of visual explanations. Unlike the nested model, which includes all necessary elements (*e.g.*, target users, domain questions) in one "situation" block, we follow the practice in current XAI frameworks [34, 42] and use four separate blocks (*Target User*, *XAI Goal*, *Usage Context*, and *Domain Explanation*) to provide clearer design guidance.

Target User describes the characteristics of users such as their AI expertise, research field, and their responsibilities in using the AI system. Previous studies [42] categorize the users into three main groups: data experts, AI novices, and AI experts. The domain users in this paper belong to the data expert group. ***XAI Goal*** relates to the motivation of explainability and clarifies which domain-related problems that the targets users aim to solve with AI explanations. It is very important to distinguish between the goal for AI and the goal for AI explanations. ***Usage Context*** depicts the context of using AI explanations (when and where), revealing characteristics such as outcome criticality, time-sensitivity, and decision complexity.

An important block that is often overlooked in previous literature is ***Domain Explanation***, which describes how a human expert would reason about a phenomenon in the applied domain. *Domain Explanation* reflects the user mental model and can help designers present AI explanations in a way that can be efficiently and accurately interpreted by users. *Domain Explanation* can vary based on *target user* and *usage context*. For example, a human expert might use inductive explanation (*e.g.*, explain using similar items) if time to make decisions is limited and use deductive explanations (*e.g.*, explain through mathematical concepts) for less time-sensitive scenarios [68].

4.3 Explanation Abstraction

This layer clarifies what explanation content and operations should be provided based on the blocks identified in domain characterization. Instead of adding an additional explanation layer to the original 4-layer nested model, we specify the data/task abstraction layer as an explanation abstraction layer by considering explanations as a special type of data. This is because most AI explanations are already described in the language of computer science, which is the fundamental purpose of using the abstraction layer in the nested model [43]. We believe such a specification is more concise and easy to use.

From literature, we identify three key blocks in the explanation abstraction layer: *Format*, *Granularity*, and *Operation*. In the original nested model, blocks are either "identified" or "designed". However, explanation abstraction should be "selected" among the possible options restricted by the existing XAI techniques. Therefore, we also enumerate the possible options for each block in the explanation abstraction layer to better guide the design of visual explanations. Meanwhile, by abstracting the three blocks, we can describe explanations in a way that is independent of the XAI algorithm details. Both ante-hoc and post-hoc explanations are supported using these abstractions. For example, rule-based explanations (an explanation format) can be generated by both ante-hoc (*e.g.*, a decision tree) or post-hoc methods (*e.g.*, deep-red [86]).

Format: Jin *et al.* [28] reviewed 59 XAI techniques and summarized three explanation formats: attribution (*e.g.*, feature importance scores), example (*e.g.*, similar examples, counterfactual examples), and clause (*e.g.*, decision trees, rule lists). This categorization is also used in later XAI frameworks [68]. However, we found that these formats, even though helpful, are difficult to be applied to summarize GNN explanations, potentially caused by the fact that the three formats are summarized from XAI techniques for Euclidean data (*e.g.*, text, image, table). The boundaries between these three formats can be vague in GNN explanations. Take node prediction in GNN as an example. Given a graph with some unlabeled nodes, a GNN predicts an unlabeled "*node m as type A*". A common explanation is that *"because node m is connected to several type A nodes"* [29, 79]. This explanation can be treated as an example-based explanation by considering other type A nodes as individual examples. On the other hand, this explanation can also be treated as attribution-based explanation by considering nodes as the attributions of the input graph.

To solve this problem and guide the visual design for GNN explanations, we conducted a review of GNN explanation techniques based on [80] and summarized three main formats for GNN explanations, *i.e.*, nodes, paths, and subgraphs. Node-based explanations show important nodes that contribute most to a certain prediction. Such explanations can be extracted from GNN models that employ the attention mechanism [66] or constructed by post-hoc methods, *e.g.*, Graph Mask [55]. Subgraph-based explanations show a subgraph of the input knowledge graph that is most related to a certain prediction. Such explanations can be extracted from GNNs that learn a local subgraph for making predictions (e.g, SEAL [85]) or constructed by post-hoc algorithms, *e.g.*, SubgraphX [81]. Path-based explanations explain a prediction through relevant paths in the knowledge graph. Such explanations can be extracted from models that consider multi-hop connections (*e.g.*, GTN [83]) or constructed by post-hoc algorithms (*e.g.*, GNN-LRP [56]).

While the explanation format is related to all the domain blocks, it is mostly influenced by the domain explanation. In other words, designers should select explanation formats that are similar to how human users explain a phenomenon to their peers [9, 63].

Granularity: granularity specifies whether to present local explanations (*i.e.*, explain individual predictions), group explanations (*i.e.*, explain a group of predictions), global explanations (*i.e.*, explain the whole model), or a combination of the above. Most existing XAI frameworks categorize explanations into local and global and rarely discuss group explanations, which is reported by our collaborators as an important level of granularity. For example, when reasoning about drug indications, domain users usually group drugs that share similar mechanism of action. A group explanation for these similar drugs can facilitate the understanding and increase the efficiency of the analysis.

Operation: Similar to the nested model, we include both low-level operations and high-level operations. A high-level ***Operation*** indicates a reasoning process users conduct upon explanations. We bring lessons from previous surveys and expert interviews [36, 42] and summarize five types of high-level operations: *why, why not, what if, how to, what else*. Other operations, such as understanding algorithms, are excluded since they are not directly related to domain-specific XAI applications, as indicated by the interview results from Sibyl [87]. Low-level operations are similar to the low-level tasks discussed in visual analytics literature [5, 76]. To accomplish high-level operations, users need to conduct a set of low-level operations such as filter explanations, compare explanations, identify abnormal explanations.

4.4 Visualization

Designers create visual encoding and interactions in this layer to present explanations to domain users, mainly driven by the three blocks in explanation abstraction. Specifically, the explanation formats should be visualized at the selected levels of granularity and the operations need to be supported through a set of interactive visualizations.

At the same time, common design practices for AI explanations should also be considered to provide familiar visualizations to users and flatten their learning curve. Some explanations are commonly represented using standard visualizations in the wild, as discussed by Wang *et al.* [68] and Jin *et al.* [28]. For example, scatter plots have been widely used to display similar and counterfactual examples; the beeswarm plot is typically used to visualize attribution explanations (*e.g.*, SHAP value) for tabular data [28, 68].

4.5 Algorithm

The Algorithm level includes the algorithmic implementation of both the interactive visualizations and the XAI techniques. An XAI algorithm should be selected and evaluated by jointly considering the output of the visualization layer, the speed of the explanation query, and the performance of the XAI algorithm (*e.g.*, stability, faithfulness). We do not distinguish ante-hoc and post-hoc explanations here since they are able to support the same explanation abstractions (*e.g.*, they both can generate local explanations). We refer readers to Vilone and Longo's survey [67] for a comprehensive list of XAI algorithms and Rubin's paper [51] for the debate about ante-hoc and post-hoc explanations.

5 DrugExplorer

This section describes how the XAI considerations introduced in Sect. 4 guide the design of DrugExplorer. Fig. 2 summarizes the design process and our evaluation strategies. We do not distinguish links within a layer and between layers for simplicity.

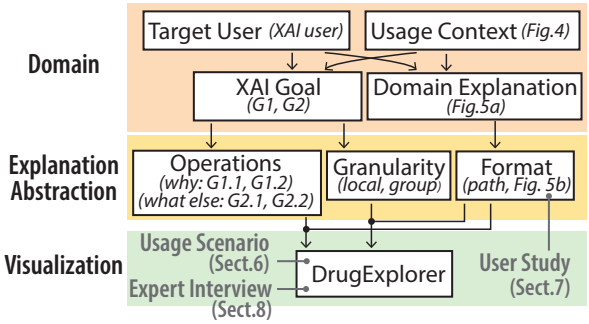

Fig. 2. Design process for DrugExplorer based on the proposed extension of the nested model with user-centric XAI considerations.

5.1 Domain

The ***Target Users*** of DrugExplorer are domain experts in drug repurposing (*e.g.*, wet lab biologists, physicians, disease experts, pharmacologists). They have limited knowledge about AI algorithms but high expertise in the application domain. As shown in Fig. 3, the typical ***Usage Context*** of DrugExplorer is after a GNN has predicted a list of drug candidates and before downstream-evaluation of these drugs. In a high-stakes task such as drug repurposing, model predictions need to be systematically evaluated by domain experts through resource-intensive laboratory experiments, including in vitro screening [21], in vivo testing [16], and clinical trials [61]. Given that only limited resources are available for such studies, domain experts need to choose a small number of highly promising therapeutic opportunities out of a number of predicted drugs. Therefore, the ***XAI Goal*** is to assist domain experts in evaluating GNN predicted drugs. Specifically, domain experts will use the explanations to **G1)** assess whether an individual drug prediction is promising and deserves further investigation; **G2)** efficiently select several most promising drugs from a potentially long list of predictions. In terms of ***Domain Explanation***, domain experts typically examine drug repurposing predictions by looking at biological processes associated with the predicted drug and reasoning how those processes relate to the disease for which the drug was predicted [21, 52]. Take Ibuprofen as an example (Fig. 4(a)). This drug can treat pain because it inhibits COX, which is required for the synthesis of prostaglandins via the arachidonic acid pathway, and prostaglandins are important mediators of pain.

5.2 Abstraction

Format: Based on the experiments on real datasets and feedback from collaborators, we can rank the three explanation formats based on their similarities to the domain explanation (Fig. 4). **A suitable explanation format for drug repurposing should mimic how a human expert explains a drug indication with biological mechanisms**. Therefore, path-based explanations are most suitable because they represent the semantic paths in the knowledge graph. For instance, the biological mechanism of Ibuprofen can be intuitively depicted by a path: *[Ibuprofen]-[COX]-[arachidonic acid pathway]-[pain]*. Explanations based on neighbor nodes are least similar to domain explanations as they mainly depict the message passing mechanism at each GNN layer. Even though the subgraph may contain some paths that make sense in the biomedical context, it can be hard for users to effectively locate these paths.

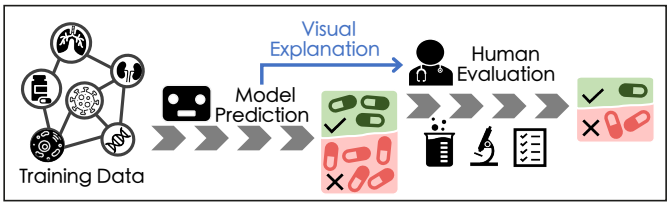

Fig. 3. DrugExplorer provides explanations to help domain experts assess drug repurposing predictions before downstream evaluation.

Operation: We selected two high-level operations, *"why"* and *"what else"*, based on the XAI goals in Sect. 5.1. The ***"why"*** operation helps users understand the reasons for a certain drug prediction (**G1**). Since the explanation for one drug can consists of multiple paths, domain users need visualizations and interactions to help them organize these explanations and generate domain insights. Specifically, users may need to **G1.1) summarize explanations based on their semantic meanings** and **G1.2) filter out less meaningful or irrelevant explanations**. The ***"what else"*** operation allows users to query similar drugs to a predicted drug for a certain disease. Grouping similar predicted drugs can accelerate the analysis of a potentially long list of drugs (**G2**). To facilitate the *"what else"* operation, we allow users to **G2.1) group similar drugs and summarize them in a domain relevant way** and **G2.2) compare different drug groups based on their explanations**. Other operations, even though promising, are excluded as they are not related to the identified domain problems. For example, the *"what if"* operation, which investigates how a modification to the input will influence the predictions, can identify new potential therapeutic opportunities by changing the structure of existing compounds. This operation is useful for drug discovery rather than drug repurposing.

Granularity: We decided granularity mainly based on the XAI goals. To support **G1**, local explanation is inevitable. For path-based explanations, local explanations can be represented as **individual paths** that correspond to how this drug perturbs the biological systems to treat a disease. Meanwhile, even though **G2** is doable by repeatedly examining local explanations, providing a group explanation for multiple similar predictions can effectively scale up the analysis. Therefore, we provide group explanations using **meta-paths**, a concept that is widely used in heterogeneous graph learning. **A meta-path is a sequence of node/edge types and can summarize paths with similar semantic meanings.** For example, the path *[Ibuprofen]-[COX]-[arachidonic acid pathway]-[pain]* belongs to the meta-path *[drug]-[protein]-[pathway]-[disease]*, which depicts a potential type of drug action mechanism.

5.3 Visualization

We designed DrugExplorer by jointly considering the three blocks of the abstraction layer. Specially, we visualize path-based explanations at different granularity levels and provide a set of interactive visualizations to support *"why"* and *"what else"* operations.

As shown in Fig. 5, DrugExplorer consists of three main components: a control panel, a drug embedding view, and an explanation view. In the control panel (a), users can search and select a disease of interest, browse the top-ranked drugs predicted by the back-end GNN model, and filter explanations through their edge importance score (**G1.2**). The drug embedding view (b) presents the learned embedding of all drugs in the knowledge graph using t-SNE [64] and highlights the predicted drugs for the selected disease. Users can easily identify similar drugs

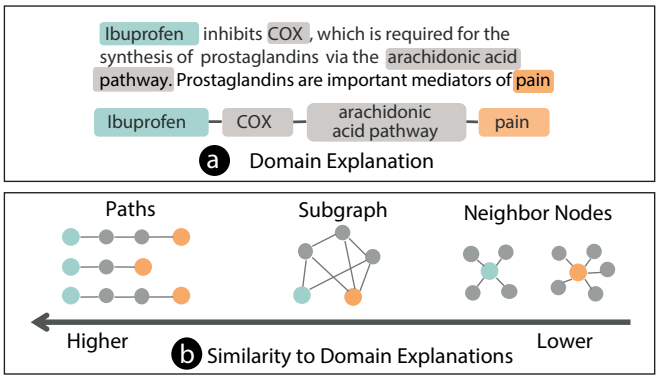

Fig. 4. (a): In the domain layer, we investigate how a domain expert would explain a drug indication. (b): In the explanation abstraction layer, we compare different GNN explanations based on their similarity to the domain explanation.

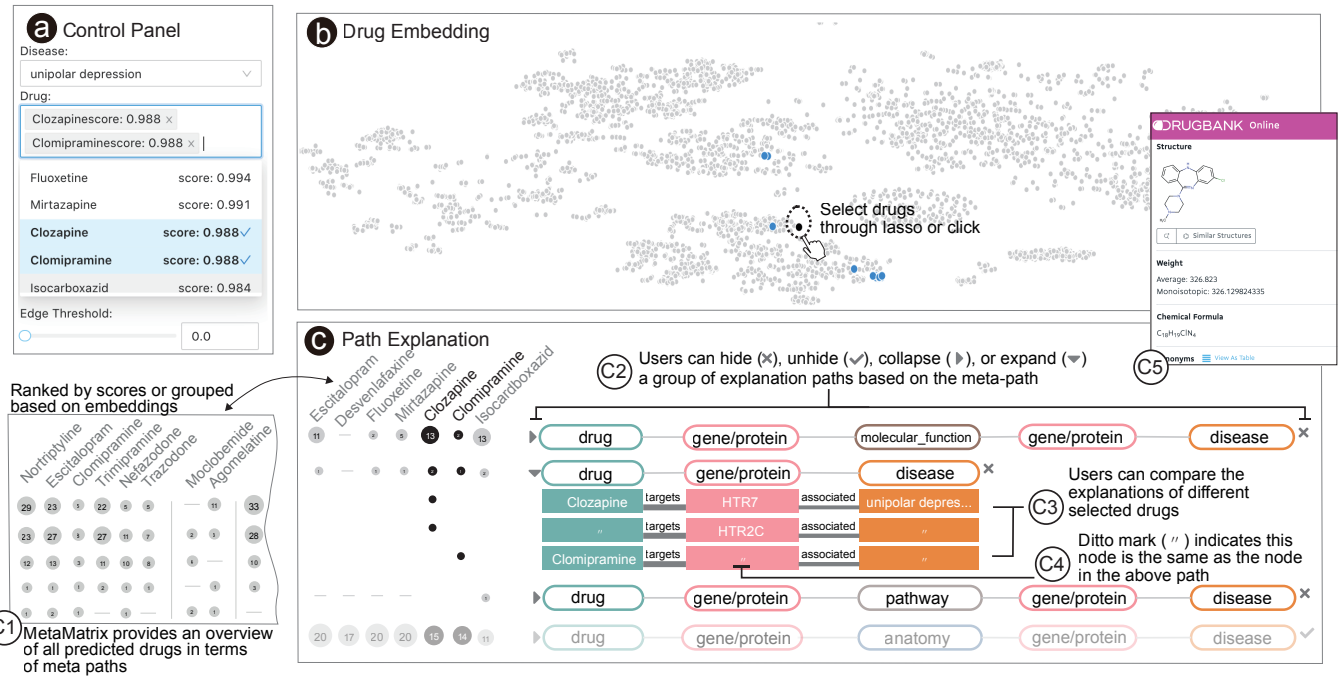

Fig. 5. DrugExplorer provides interactive visual explanations for GNN-based drug repurposing. Users can select drugs based on their rankings using the control panel (a) or their similarities using the drug embedding view (b). The explanation view (c) incorporates a novel *MetaMatrix* design and provides diverse interactions (C1-5) for users to effectively interpret and validate explanations.

in this embedding space (**G2.1**). The explanation view (c) provides path-based explanations for individual drug predictions.

The explanation view incorporates a novel *MetaMatrix* design. This design is inspired by the matrix design in [72] to enable user examine, summarize, and compare explanations at different granularity. In *MetaMatrix*, each column is a predicted drug; each row is a meta-path, which can be expanded to rows of the corresponding paths. Meta-path is a concept widely used in heterogeneous graph learning. It uses a sequence of node/sequence types to summarize paths. We use different encodings to distinguish meta paths and paths, *i.e.*, nodes in meta-paths are represented as rounded rectangles with borders while nodes in paths are represented as rectangles with solid fills. The number in each cell indicates the number of explanation paths that belong to the corresponding meta-path or path (**G1.1**).

MetaMatrix provide diverse user interactions. First, the drugs (*i.e.*, columns) can be sorted based on their prediction scores or grouped based on their proximity in the embedding space (**G2.1**). Users can efficiently compare different drugs (individual columns) or different groups (grouped columns) of drugs in terms of meta-paths, including the length of meta-paths, node types in meta-paths, and the number of paths belonging to a meta-path (**G2.2**) of interest. Second, users can hide, un-hide explanation paths (C2) to focus on the explanations of interest (**G1.2**). For example, as shown in Fig. 5, users can collapse other meta-paths to highlight the comparison on interesting meta-paths (*i.e.*, *[disease]-[protein]-[drug]*, *[disease]-[protein]-[phenotype]-[protein]-[drug]*). Users can also hide explanations of a specific meta-path if they think the related mechanism is less convincing. For instance, the meta-path that shows the drug protein and the disease protein are both absent in the same anatomy (*i.e.*, *[disease]-[protein]-[anatomy]-[protein]-[drug]*) is less convincing than the explanation that the drug protein and the disease protein are connected to the same pathway (*i.e.*, *[disease]-[protein]-[pathway]-[protein]-[drug]*). Third, users can expand meta-paths and compare drugs on a more detailed level based on individual explanation paths. For example, as shown in Fig. 5(C3), *Clozapine* and *Clomipramine* are predicted for treating the disease *unipolar depression* partly because they are both connected to *HTR2C*, a protein that is connected to *unipolar depression*. *Clozapine* is at the left side and has a higher rank than *Clomipramine*, which might be related to the fact that *Clozapine* is also connected to another protein *HTR7*. Meanwhile, to help users quickly identify similar explanation paths, we employ the ditto mark (") (C4) to indicate that a node has the same name as the node in the path above (**G1.1**). Users can also review drug details from the DrugBank database [77] in a pop-up window (C5).

5.4 Algorithm

Training Datasets. The training data for our study is a heterogeneous knowledge graph consisting of 10 different types of entities (*e.g.*, drug, disease, protein) and 32 semantically distinct types of relationships between the entities (*e.g.*, drug-disease indications, protein-protein interactions, drug-protein interactions). The dataset was assembled from 21 public databases of protein-protein interactions, gene expression data, clinical trials, and drug usage across the entire range of 22K+ human diseases and 7K+ drugs.

GNN Model and Explanations. We formulated drug repurposing as a link prediction task. The GNN model tries to predict among three link types $r \in \mathscr{R}$ (*i.e.*, indication, contra-indication, or off-label use) between a drug and a disease that are not connected in the training data (*i.e.*, their relationship is unknown).

We used a heterogeneous GNN to generate embeddings for every node in the knowledge graph. Specifically, for a node i at the GNN layer l, its embedding $\mathbf{h}_i^{(l)}$ is calculated by aggregating the embeddings from the previous layer of its neighbor nodes $\mathscr{N}_i$, using relation weight matrices $\mathbf{W}_r^{(l)}$ and a message calculation function f: $\mathbf{h}_i^{(l)} = \mathbf{h}_i^{(l-1)} + \sum_{r\in\mathscr{R}} \sum_{j\in\mathscr{N}_i} f(\mathbf{W}_r^{(l)}, \mathbf{h}_j^{(l-1)})$. Given the embedding of a drug i and a disease j, we predict the probability of edge relation r as $p_{i,j,r} = 1/(1+\exp(-\text{sum}(\mathbf{h}_i * \mathbf{w}_r * \mathbf{h}_j)))$. We show that this model can accurately predict drug-disease relationships: the predicted drug-disease relationships rank 79.5% of hits in the top 5%, and 88.9% of hits in the top 10%.

To provide high-quality path-based explanations at both group and local level, we experimented with and adapted different ante-hoc and post-hoc explanation methods, including Graph Attention [74], GNNExplainer [79], and GraphMask [55]. We selected GraphMask due to its high fidelity. Finally, we developed a post-hoc graph explainability based on GraphMask that can drop superfluous edges from the knowledge graph and only retain a sparse set of edges that contribute most towards the prediction (Supplementary Sect.S3).

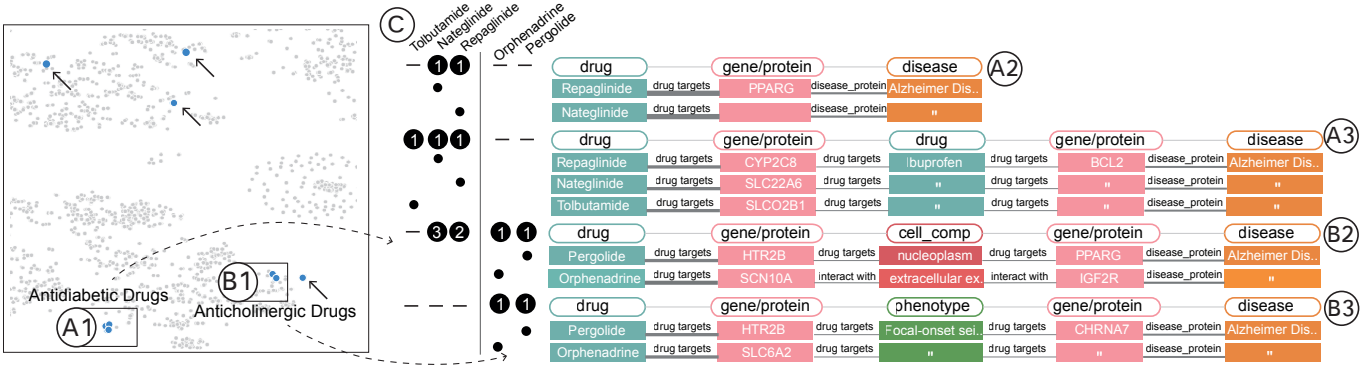

Fig. 6. Exploring two groups of drug repurposing predictions (A1, B1) for the treatment of Alzheimer's Disease.

5.5 Implementation

The interactive visual explanations are implemented in JavaScript using React.js [26], D3.js [7], and Ant Design [62]. The GNN model is implemented in Python using Pytorch [50]. The graph data is stored in Neo4j database [44]. The visual explanations communicate with the back-end GNN model through a Python web server built with Flask [20]. The source code and an interactive demo are available at `https://github.com/hms-dbmi/Drug_Explorer`.

6 Usage Scenario

We demonstrate how DrugExplorer can be used to examine treatments for Alzheimer's Disease (AD). The GNN model was trained on the full knowledge graph and used to make predictions for drugs that were not included in the knowledge graph. We selected AD in the visualization tool and explored predicted drugs and their explanations.

The tool automatically produced predictions, explanations and updated visualizations for AD (Fig. 6). Predicted drugs were scattered in the *Embedding* view, indicating that the GNN model produced predictions for a diverse set of drugs.

We first examined the largest cluster of drugs (Fig. 6(A1)). This cluster included drugs such as *Glyburide*, *Repaglinide*, *Tolbutamide*, and *Metformin*, which are commonly used to treat Type 2 diabetes (T2D). Drugs found in the cluster were consistent with the current scientific understanding of the connections between cognitive impairment and T2D [53]. Previous studies found that the use of antidiabetic treatments among individuals with T2D could mitigate risk for dementia [3].

We then examined explanations for the predicted antidiabetic drugs in the *MetaMatrix* view. To this end, we first selected *Repaglinide* in the *MetaMatrix* view to show detailed explanations. The shortest meta-path is *Disease-Gene/Protein-Drug*. The explanation path below that meta-path (Fig. 6.A2) showed that *Repaglinide* targets protein *PPARG*, which, in turn, is associated with AD. Based *Disease-Gene /Protein-Drug-Gene/Protein-Disease* meta-path (A3), we see that drug *Repaglinide* was predicted partly because it has the same target protein as *Ibuprofen*. *Ibuprofen* targets proteins that are associated with AD and can delay some forms of AD pathology [37]. Similar instances of meta-paths existed in explanations of other antidiabetic drugs, including *Nateglinide* and *Tolbutamide* (A2, A3).

Another cluster (Fig. 6(B1)) in the *Embedding* view comprised of anticholinergic drugs, including *Pergolide* and *Orphenadrine*, which are used to manage Parkinson's disease. Based on the *MetaMatrix*(C), we found this drug group is different from the previous T2D group in terms of meta-paths. Specifically, the explanations for this group did not have *Disease-Gene/Protein-Drug* or *Disease-Gene /Protein-Drug-Gene/Protein-Disease*, which were the main explanations for T2D drug group. We then investigated the explanation paths for more details. We found that the target protein of *Pergolide* and *Orphenadrine* interacts with multiple AD-associated proteins through shared cellular phenotypes (B2), an observation consistent with the reported associations between AD and anti-Parkinson's agents [46]. While some studies [30] reported the contraindication of these drugs, the contraindication still reflected the GNN's ability to identify associations unknown in the training graph. This example also highlighted the utility of visual explanations to involve humans and identify possible inaccurate predictions.

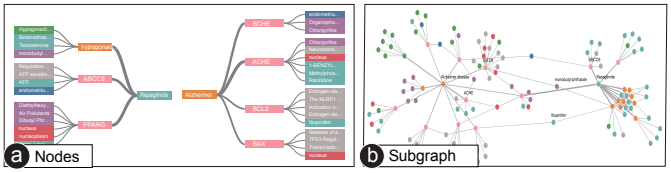

Fig. 7. The user study compares path-based explanations against three alternative conditions: node-based explanations (a), subgraph-based explanations (b), and an non-explanation baseline.

7 User Study

Participants. We recruited 12 medical professionals (7 males, 5 females, denoted as P1-12) through personal contacts, Slack channels, and email lists in related institutions. The mean (SD) age of the participants was 34.25 (6.12) years. All participants have worked in medicine-related fields for more than five years, including five clinical researchers (P1-3, P11-12) and five practicing physicians (P4, P7-10), who all have MD degrees, and two medical school students who used to work as pharmacists (P5, P6). The participants were familiar with basic concepts of machine learning but are not experts. No participants knew this project before and none of them are authors of this paper.

Conditions and Tasks. We tested total four conditions: 1) a node-based explanation; 2) a path-based explanation; 3) a subgraph-based explanation; and 4) an non-explanation baseline that only reported a confidence score. Since we aim to assess the visual presentations independent from the algorithmic aspect of explanations, we used the same algorithm (*i.e.*, GraphMask [55]) and generated explanations with different presentations through certain transformations (Supplementary Material, Sect. S1.4). For all the three visualizations, the color indicates the node type, and edge line-width indicates the importance. Users can interactively filter explanations based on their importance.

We collected 16 predicted drug-disease treatment pairs (twelve correct, four wrong) and asked the participants to assess these predictions under four different conditions (four predictions in each condition). Since other alternatives can not effectively group explanations, we only asked users to evaluate individual predictions.

The tasks and the evaluation procedure were validated and refined through a pilot study with two domain experts and one AI expert. The three pilot study participants were not included in the twelve participants of the study reported here. The two domain experts were not authors but the AI expert is an author of this paper. The full list of the drug-disease pairs and the interface used for the user study are described in the Supplementary Material.

Procedure. The evaluation took around 40 minutes on average for each participant. Participants were first presented with a brief introduction about the study, an informed consent form, and a 10-min tutorial

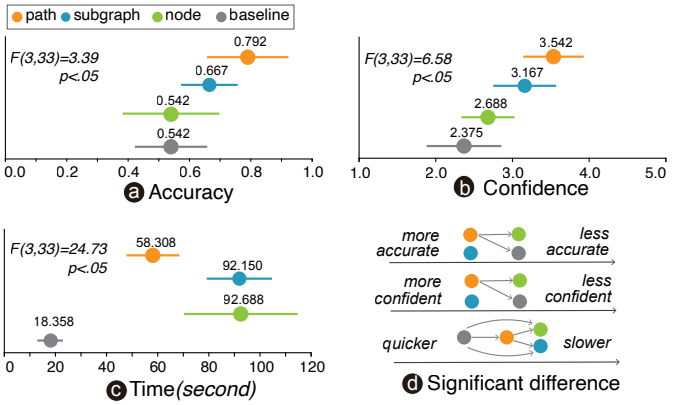

Fig. 8. Accuracy (a), confidence levels (b), and time (c) under different conditions. Error bars indicate 95% confidence intervals. A link in (d) indicates a significant difference between two conditions.

about how to read and interact with the three visual explanations. Participants assessed 16 AI predictions under four different conditions (four predictions under each condition). For each prediction, participants decided whether the predicted drug can be used for treating a certain disease and reported their confidence levels using a 5-point Likert scale (1=not confident at all, 5=completely confident). The completion time for assessing each prediction was automatically recorded by our study system. The order of predictions and the order of the four conditions were randomized and counterbalanced across participants. Finally, we asked the participants a set of semi-structured questions around two main topics: 1) which factors influenced their decisions and their confidence level; 2) how they interpreted the AI explanations.

Results. We set $\alpha = 0.05$ and tested three hypotheses: **H1)** Path-based explanations have higher accuracy than other conditions; **H2)** Path-based explanations enable more confident user performance than other conditions; **H3)** Path-based explanations require less time than other explanation types but more time than baseline. Another purpose of assessing the 16 predictions is to force participants to actually make decisions using different explanations, which are important to precisely understand user perception of AI explanations and generate helpful discussions in the following interviews [8]. We conducted the Repeated Measures ANOVA analysis to compare the average accuracy, self-reported confidence score, and completion time across the four conditions. If there is a significant difference among the four conditions, we ran Tukey's Honest Significant Difference test to confirm whether the differences occurred between each two conditions.

Results of the user study are summarized in Fig. 8. Path-based explanations have significantly better performance than baseline and node-based explanations at all three metrics: accuracy, confidence, time. Compared with subgraph-based explanation, even though the path-based explanations' advantages are not significant in terms of accuracy and confidence, they require significantly less time. Surprisingly, the user study results show that providing explanations does not necessarily improve user performance. Node-based explanations and subgraph-based explanations do not have significantly higher accuracy or confidence than baseline. Participants' ratings for the three types of visual explanations were roughly consistent with their similarity to domain explanations, as shown in Fig. 4(b).

8 Expert Interview

Ten out of the twelve participants in the user study agreed to participate in an interview about their experience with DrugExplorer. During the interview, we first demonstrated the functionalities of DrugExplorer using the usage scenario about Alzheimer's Disease Sect. 6. Participants then freely explored the diseases and drugs of interest on a testing set containing 48 diseases (Supplementary Material, Sect. S1). Each participant selected at least one disease of interest, explored the interactive visualizations, and freely commented on the AI predictions, the visual explanations, and their usage experiences of DrugExplorer. The interview took around 25 minutes for each participant.

Overall, participants expressed great interests in this tool, commented that it *"targets an important problem and can be super helpful"*. Even though we introduced a new visualization design, *MetaMatrix*, all participants agreed that they had no difficulties in understanding the AI explanations and interacting with the tool. Meanwhile, participants exhibited cautious enthusiasm towards DrugExplorer and emphasized that downstream evaluations, such as clinical trials, were essential to validate the AI-predicted drug repurposing, even if only for regulatory purposes. For example, P8, a physician who specializes in pain management and *"prescribed a lot of off-label drugs"*, expressed strong interest in using this tool since the explanations were consistent with his reasons for some off-label prescriptions. They described his plan for validating the potential drug candidates: i) identify promising drug candidates whose explanations are biomedically meaningful; ii) validate the biomedical mechanisms in the explanation and ensure the drug has no adverse effect; iii) prescribe this drug to some patients who are not responding to first-line treatments (*i.e.*, approved or recommended treatments); iv) conduct clinical trials if the drug seems effective.

Most participants agreed that DrugExplorer supports the goal of repurposing drugs well. In addition, some participants commented that this tool can potentially be generalized beyond drug repurposing to other related problems. For example, P4 commented that this tool could serve as an educational tool to help medical students better understand existing drugs, diseases, and their relationships to other medical entities. P1 and P2 stated that this tool could be used for polypharmacy (*i.e.*, the simultaneous use of multiple drugs), such as predicting polypharmacy side effects. *"Similar to explaining a drug-disease indication, the visualization can show how a drug changes the activities of another drug and illustrates the causes of side effects."* (P2).

Participants also offered helpful suggestions for improving this tool. Five participants mentioned that more biomedical information about the nodes and edges would help them more confidently assess the explanations. *"They [the provided explanations] are useful but somehow abstract."* (P8) *"The [disease] - [protein] - [pathway] - [protein] - [disease] can be a strong evidence but I need to know more details about how this protein is involved in this pathway. I can always check literature for such information myself, but it would be great if it is provided here."* (P2) P10 suggested the functionality to annotate explanations, save, and share these annotations. These suggestions reflect the participants' wishes to better align the AI explanations with how they typically reason about a drug indication, indicating the importance of choosing proper explanation abstractions based on domain characterizations. Meanwhile, three participants (P2, P5, P9) mentioned that the subtle distinctions between the represented explanations and real-world biomedical mechanism can be sometimes confusing. For example, in AI explanations, edge thickness represents the importance of this edge to a certain AI prediction. The thickness can be easily confused with the strength of the biological relation.

9 Observations, Insights, and Discussion

9.1 Observations about domain users

Human Knowledge vs. AI Explanation. We did not observe blind trust in AI explanations as reported in some previous studies [31, 68], which might be related to the critical nature of the medical domain. Instead, participants heavily relied on their prior knowledge accumulated through years of experience and medical training when assessing predictions and explanations. All participants stated they first used their own knowledge when checking the predictions. When prior knowledge could help them make a decision, most (9/12) participants stated that they still examined explanations to validate their decisions and evaluate the quality of explanations. When participants were not familiar with the drug or the disease, they examined whether the AI explanation is domain relevant. For example, some (5) participants said that *[disease] - [protein] - [drug]* was strong evidence, because this path indicated that the disease is directly associated with the drug's target protein. On the contrary, *[disease] - [protein] - [anatomy] - [protein] - [drug] "is more like a correlation rather than a causation"* (P2).

Domain explanations can vary slightly across human experts. While participants employed similar ways of reasoning about a drug indication (*i.e.*, checking the connections between the drug and the disease), we also observed subtle differences among participants. For example, P5 stated that they *"consider[ed] the drug and the disease simultaneously to see how they met in the middle"*. The path-based explanations in DrugExplorer were confusing at first because P5 felt they need to read from left (drug) to right (disease). But P5 also commented that this problem is *"easy to overcome after exploring some predictions"*.

Actually making decisions influence human experts' opinions towards explanations. We observed that the attitude of some participants (P2, P4, P5) towards the three explanation types changed before and after assessing the 16 predictions. This indicates the importance of interacting with AI explanations and performing actual tasks in evaluating XAI. For example, P2 commented *"the subgraph one is so much better than others"* when learning the tutorial. However, in the post-study interview, P2 stated *"this [path-based] explanation can provide all the information I needed when checking that [subgraph-based] explanation, and even in a more straightforward way. The subgraph is just more visually appealing to me."* In earlier studies, expert interviews that were purely based on imaginary scenarios or non-explainable AI predictions are widely used [9, 63]. They provide an efficient approach to understand user needs and preferences, especially considering the numerous time and efforts required to develop XAI systems. However, our observations suggest that, without interacting with AI explanations and completing actual tasks, participants may report inaccurate feedback in some situations.

Human experts tend to reshape less suitable explanations. All participants stated that examining the connections between the drug and disease is their primary way of assessing an AI prediction. When using node-based explanations, which is inconsistent with their preferred reasoning processes, participants responded differently. Two participants still tried to find connections by identifying same nodes in the neighbors of the drug and the neighbors of the diseases, which are *"extremely painful to find useful information"* (P3). Other participants, however, treated the node-based explanations as providing context information about the drug and the disease. For example, P8 mentioned that they mainly checked whether the *phenotype* and *pathway* nodes in the drug's neighbors were related to the disease based on his prior knowledge. In other words, participants tried to build an implicit connection between the drug and the disease using their prior knowledge. One possibility is that, when the explanation is too far from their mental models, participants tended to reshape the explanation and add extra information to make the explanation similar to their mental models. While this strategy made node-based explanations easier to interpret, participants were less confident about their interpretation, which is reflected in their reported confidence levels (Fig. 8(b)).

9.2 Reflections on XAI visual design

Contextualized XAI through visualization design. Current XAI algorithms usually construct explanations via a data-centric approach regardless of the context in which the explanation is used. While such a strategy ensures these algorithms are generic and applicable to a variety of problems, it also poses challenges for human experts in interpreting these explanations and obtaining actionable insights. Our study suggests that the one-size-fits-all explanations can fail in real-world applications. When facing explanations that are inconsistent with their commonly-used domain explanations, users tend to reshape these explanations to match their mental models and become less confident about their interpretation. In spite of the importance of context, it can be challenging to integrate context factors into XAI algorithms.

Our study shows that visualization designs can be considered independently from the algorithmic aspects and serve as an effective method to integrate the context of an explanation. DrugExplorer is designed and developed guided by a list of context factors (*i.e.*, the domain explanations, the usage contexts, the XAI goal, target users) that we identified through literature review and collaborators' feedback. Meanwhile, this list of context factors is not exhaustive and will evolve as future design studies and field studies are conducted. For example, *Domain Explanation*, an important design consideration revealed in our design study, is only briefly discussed in previous studies. We anticipate that this paper will encourage more design studies and field studies to better understand how to contextualize XAI through visualization design.

Interactive visualization is a fundamental component of AI explanations. This study suggests that, apart from algorithms that extract information for explaining a prediction, visual presentation and user interaction are also critical components of an AI explanation, especially in human-AI collaboration. As we demonstrate, how an AI explanation is visually presented and how users interact with the explanation can directly influence how users interpret and use the explanation.

More importantly, providing AI explanations with proper interactive visualizations not only helps users interpret the explanations but also encourages feedback from users. For example, our study participants employed the *hide* interaction to hide meta-paths that are not meaningful in the biomedical context. Such interactions reflect users' domain knowledge and act as important feedback. In future work, we plan to integrate such feedback into the model training, which can improve the performance of the AI and the quality of explanation.

9.3 Limitations and future work

While the evaluation demonstrates the effectiveness of DrugExplorer and the extend nested model, this study has several limitations. First, we conducted the evaluation in the setting of a laboratory study rather than in a real-world deployment. This limitation is shared with many other prototype visualization tools [22, 47, 70]. More importantly, a real-world deployment of DrugExplorer can be challenging due to the regulatory and ethical issues involved in drug repurposing. At the same time, participants reported positive feedback about DrugExplorer and agreed with its usability. The evaluation generated valuable observations and findings that will benefit future applications of XAI. Second, limited by the training data and the back-end GNN model, the explanation format is relatively simple and may not provide all detailed information a human expert needs to systematically assess a drug repurposing prediction. For example, for edges in the knowledge graph, the back-end GNN model only considers edge types. Therefore, DrugExplorer does not provide edge details such as the protein binding sites targeted in a *[drug]-[protein]* edge. Even though the GNN model already generates accurate predictions and explanations, providing more biomedical details can better assist human experts. We plan to further improve the knowledge graph and incorporate detailed biomedical information in the visual explanations in future work.

10 Conclusion

This paper presents a design study that investigates how to select and visualize AI explanations for domain experts in GNN-based drug repurposing. This design study follows the nested model of visualization design and extends it by incorporating user-centric XAI considerations based on a literature review and feedback from collaborators. An interactive visualization tool, DrugExplorer, is designed, developed, and evaluated. DrugExplorer provides a novel visualization called *MetaMatrix* that enables efficient organization and comparison of explanation paths at different granularity. This design can be applied to other similar problems, such as explaining GNN-predicted polypharmacy side effects. Our extension to the nested model highlights important takeaways: (1) visualization of explanations should consider both the domain users' mental model and the available explanation formats; (2) the needed interactions are related to the XAI goals as well as the supported XAI operations by existing techniques. This extension does not aim to be an exhaustive list, but a cornerstone that will inspire and be further extended through future design studies and field studies.

Acknowledgments

The authors wish to thank all the participants in the expert interviews and user studies. M.Z. is supported, in part, by NSF under Nos. IIS-2030459 and IIS-2033384, Air Force Contract No. FA8702-15-D-0001, Harvard Data Science Initiative, Amazon Research Award, Bayer Early Excellence in Science Award, AstraZeneca Research, and Roche Alliance with Distinguished Scientists Award.

REFERENCES

[1] J. Adebayo, M. Muelly, I. Liccardi, and B. Kim. Debugging tests for model explanations. In *Advances in Neural Information Processing Systems*, vol. 33, pp. 700–712. Curran Associates, Inc., 2020.

[2] C. Agarwal, M. Zitnik, and H. Lakkaraju. Towards a rigorous theoretical analysis and evaluation of gnn explanations. *arXiv preprint arXiv:2106.09078*, 2021.

[3] H. Akimoto, A. Negishi, S. Oshima, H. Wakiyama, M. Okita, N. Horii, N. Inoue, S. Ohshima, and D. Kobayashi. Antidiabetic drugs for the risk of alzheimer disease in patients with type 2 dm using faers. *American Journal of Alzheimer's Disease & Other Dementias®*, 35:1533317519899546, 2020.

[4] A. Alqaraawi, M. Schuessler, P. Weiß, E. Costanza, and N. Berthouze. Evaluating saliency map explanations for convolutional neural networks: a user study. In *Proceedings of the 25th International Conference on Intelligent User Interfaces*, pp. 275–285, 2020.

[5] R. Amar, J. Eagan, and J. Stasko. Low-level components of analytic activity in information visualization. In *IEEE Symposium on Information Visualization, 2005. INFOVIS 2005.*, pp. 111–117. IEEE, 2005.

[6] O. Anuyah, W. Fine, and R. Metoyer. Design decision framework for ai explanations. In C. Wienrich, P. Wintersberger, and B. Weyers, eds., *Mensch und Computer 2021 - Workshopband*. Gesellschaft für Informatik e.V., Bonn, 2021. doi: 10.18420/muc2021-mci-ws02-237

[7] M. Bostock, V. Ogievetsky, and J. Heer. D^3 data-driven documents. *IEEE Transactions on Visualization and Computer Graphics*, 17(12):2301–2309, 2011.

[8] Z. Buçinca, P. Lin, K. Z. Gajos, and E. L. Glassman. Proxy tasks and subjective measures can be misleading in evaluating explainable ai systems. In *Proceedings of the 25th International Conference on Intelligent User Interfaces*, pp. 454–464, 2020.

[9] C. J. Cai, S. Winter, D. Steiner, L. Wilcox, and M. Terry. "hello ai": Uncovering the onboarding needs of medical practitioners for human-ai collaborative decision-making. *Proceedings of the ACM on Human-computer Interaction*, 3(CSCW):1–24, 2019.

[10] K. Cao, M. Liu, H. Su, J. Wu, J. Zhu, and S. Liu. Analyzing the noise robustness of deep neural networks. *IEEE Transactions on Visualization and Computer Graphics*, 27(7):3289–3304, 2021.

[11] S. K. Card, J. D. Mackinlay, and B. Shneiderman. Readings in information visualization: using vision to think, 1999.

[12] S. Chari, O. Seneviratne, D. M. Gruen, M. A. Foreman, A. K. Das, and D. L. McGuinness. Explanation ontology: A model of explanations for user-centered ai. In *International Semantic Web Conference*, pp. 228–243. Springer, 2020.

[13] C. Chen, J. Yuan, Y. Lu, Y. Liu, H. Su, S. Yuan, and S. Liu. OoDAnalyzer: Interactive analysis of out-of-distribution samples. *IEEE Transactions on Visualization and Computer Graphics*, 27(7):3335–3349, 2021.

[14] F. Cheng, D. Liu, F. Du, Y. Lin, A. Zytek, H. Li, H. Qu, and K. Veeramachaneni. Vbridge: Connecting the dots between features and data to explain healthcare models. *IEEE Transactions on Visualization and Computer Graphics*, 2021.

[15] H.-F. Cheng, R. Wang, Z. Zhang, F. O'Connell, T. Gray, F. M. Harper, and H. Zhu. Explaining decision-making algorithms through ui: Strategies to help non-expert stakeholders. In *Proceedings of the 2019 chi conference on human factors in computing systems*, pp. 1–12, 2019.

[16] R. Diaz-Gonzalez, F. M. Kuhlmann, C. Galan-Rodriguez, L. M. da Silva, M. Saldivia, C. E. Karver, A. Rodriguez, S. M. Beverley, M. Navarro, and M. P. Pollastri. The susceptibility of trypanosomatid pathogens to PI3/mTOR kinase inhibitors affords a new opportunity for drug repurposing. *PLoS Neglected Tropical Diseases*, 5(8):e1297, 2011.

[17] M. Eiband, H. Schneider, M. Bilandzic, J. Fazekas-Con, M. Haug, and H. Hussmann. Bringing transparency design into practice. In *23rd international conference on intelligent user interfaces*, pp. 211–223, 2018.

[18] S. Feng and J. Boyd-Graber. What can AI do for me? evaluating machine learning interpretations in cooperative play. In *Proceedings of the 24th International Conference on Intelligent User Interfaces*, pp. 229–239, 2019.

[19] J. J. Ferreira and M. S. Monteiro. What are people doing about xai user experience? a survey on ai explainability research and practice. In *International Conference on Human-Computer Interaction*, pp. 56–73. Springer, 2020.

[20] M. Grinberg. *Flask web development: developing web applications with python*. " O'Reilly Media, Inc.", 2018.

[21] D. M. Gysi, Í. Do Valle, M. Zitnik, A. Ameli, X. Gan, O. Varol, S. D. Ghiassian, J. Patten, R. A. Davey, J. Loscalzo, et al. Network medicine framework for identifying drug-repurposing opportunities for covid-19. *Proceedings of the National Academy of Sciences*, 118(19), 2021.

[22] T. A. Harbig, S. Nusrat, T. Mazor, Q. Wang, A. Thomson, H. Bitter, E. Cerami, and N. Gehlenborg. Oncothreads: visualization of large-scale longitudinal cancer molecular data. *Bioinformatics*, 37(Supplement_1):i59–i66, 2021.

[23] F. Hohman, A. Head, R. Caruana, R. DeLine, and S. M. Drucker. Gamut: A design probe to understand how data scientists understand machine learning models. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*, pp. 1–13, 2019.

[24] S. R. Hong, J. Hullman, and E. Bertini. Human factors in model interpretability: Industry practices, challenges, and needs. *Proceedings of the ACM on Human-Computer Interaction*, 4(CSCW1):1–26, 2020.

[25] K. Huang, T. Fu, W. Gao, Y. Zhao, Y. Roohani, J. Leskovec, C. W. Coley, C. Xiao, J. Sun, and M. Zitnik. Therapeutics data commons: machine learning datasets and tasks for therapeutics. *arXiv preprint arXiv:2102.09548*, 2021.

[26] F. Inc. React.js. `https://github.com/facebook/react`.

[27] J. Jiménez-Luna, F. Grisoni, and G. Schneider. Drug discovery with explainable artificial intelligence. *Nature Machine Intelligence*, 2(10):573–584, 2020.

[28] W. Jin, S. Carpendale, G. Hamarneh, and D. Gromala. Bridging ai developers and end users: An end-user-centred explainable ai taxonomy and visual vocabularies. *Proceedings of the IEEE Visualization, Vancouver, BC, Canada*, pp. 20–25, 2019.

[29] Z. Jin, Y. Wang, Q. Wang, Y. Ming, T. Ma, and H. Qu. Gnnlens: A visual analytics approach for prediction error diagnosis of graph neural networks. *arXiv preprint arXiv:2011.11048*, 2020.

[30] K.-i. Joung, S. Kim, Y. H. Cho, and S.-i. Cho. Association of anticholinergic use with incidence of alzheimer's disease: population-based cohort study. *Scientific reports*, 9(1):1–10, 2019.

[31] H. Kaur, H. Nori, S. Jenkins, R. Caruana, H. Wallach, and J. Wortman Vaughan. Interpreting interpretability: Understanding data scientists' use of interpretability tools for machine learning. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*, pp. 1–14, 2020.

[32] S. Konecni, J. Zhou, and G. Grinstein. A visual analytics model applied to lead generation library design in drug discovery. In *2009 13th International Conference Information Visualisation*, pp. 345–352. IEEE, 2009.

[33] B. C. Kwon, M.-J. Choi, J. T. Kim, E. Choi, Y. B. Kim, S. Kwon, J. Sun, and J. Choo. Retainvis: Visual analytics with interpretable and interactive recurrent neural networks on electronic medical records. *IEEE Transactions on Visualization and Computer Graphics*, 25(1):299–309, 2018.

[34] M. Langer, D. Oster, T. Speith, H. Hermanns, L. Kästner, E. Schmidt, A. Sesing, and K. Baum. What do we want from explainable artificial intelligence (xai)?–a stakeholder perspective on xai and a conceptual model guiding interdisciplinary xai research. *Artificial Intelligence*, 296:103473, 2021.

[35] A. Lex, C. Partl, D. Kalkofen, M. Streit, S. Gratzl, A. M. Wassermann, D. Schmalstieg, and H. Pfister. Entourage: Visualizing relationships between biological pathways using contextual subsets. *IEEE Transactions on Visualization and Computer Graphics*, 19(12):2536–2545, 2013.

[36] Q. V. Liao, D. Gruen, and S. Miller. Questioning the AI: informing design practices for explainable ai user experiences. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*, pp. 1–15, 2020.

[37] G. P. Lim, F. Yang, T. Chu, P. Chen, W. Beech, B. Teter, T. Tran, O. Ubeda, K. H. Ashe, S. Frautschy, et al. Ibuprofen suppresses plaque pathology and inflammation in a mouse model for alzheimer's disease. *Journal of Neuroscience*, 20(15):5709–5714, 2000.

[38] S. M. Lundberg, B. Nair, M. S. Vavilala, M. Horibe, M. J. Eisses, T. Adams, D. E. Liston, D. K.-W. Low, S.-F. Newman, J. Kim, et al. Explainable machine-learning predictions for the prevention of hypoxaemia during surgery. *Nature biomedical engineering*, 2(10):749–760, 2018.

[39] M. Meyer, M. Sedlmair, P. S. Quinan, and T. Munzner. The nested blocks and guidelines model. *Information Visualization*, 14(3):234–249, 2015.

[40] T. Miller. Explanation in artificial intelligence: Insights from the social sciences. *Artificial Intelligence*, 267:1–38, 2019.

[41] Y. Ming, H. Qu, and E. Bertini. Rulematrix: Visualizing and understanding classifiers with rules. *IEEE Transactions on Visualization and Computer Graphics*, 25(1):342–352, 2018.

[42] S. Mohseni, N. Zarei, and E. D. Ragan. A multidisciplinary survey and framework for design and evaluation of explainable ai systems. *ACM Transactions on Interactive Intelligent Systems (TiiS)*, 11(3-4):1–45, 2021.

[43] T. Munzner. A nested model for visualization design and validation. *IEEE Transactions on Visualization and Computer Graphics*, 15(6):921–928, 2009.

[44] I. Neo4j. Neo4j graph data platform. `https://neo4j.com`. accessed: 2020-10-01.

[45] N. Nosengo. New tricks for old drugs. *Nature*, 534(7607):314–317, 2016.

[46] K. Ono, K. Hasegawa, H. Naiki, and M. Yamada. Anti-parkinsonian agents have anti-amyloidogenic activity for alzheimer's β-amyloid fibrils in vitro. *Neurochemistry International*, 48(4):275–285, 2006.

[47] C. Partl, S. Gratzl, M. Streit, A. M. Wassermann, H. Pfister, D. Schmalstieg, and A. Lex. Pathfinder: Visual analysis of paths in graphs. In *Computer Graphics Forum*, vol. 35, pp. 71–80. Wiley Online Library, 2016.

[48] C. Partl, A. Lex, M. Streit, D. Kalkofen, K. Kashofer, and D. Schmalstieg. enroute: Dynamic path extraction from biological pathway maps for exploring heterogeneous experimental datasets. *BMC Bioinformatics*, 14(19):1–16, 2013.

[49] C. Partl, A. Lex, M. Streit, H. Strobelt, A.-M. Wassermann, H. Pfister, and D. Schmalstieg. Contour: data-driven exploration of multi-relational datasets for drug discovery. *IEEE Transactions on Visualization and Computer Graphics*, 20(12):1883–1892, 2014.

[50] A. Paszke, S. Gross, F. Massa, A. Lerer, J. Bradbury, G. Chanan, T. Killeen, Z. Lin, N. Gimelshein, L. Antiga, A. Desmaison, A. Kopf, E. Yang, Z. DeVito, M. Raison, A. Tejani, S. Chilamkurthy, B. Steiner, L. Fang, J. Bai, and S. Chintala. Pytorch: An imperative style, high-performance deep learning library. In H. Wallach, H. Larochelle, A. Beygelzimer, F. d'Alché-Buc, E. Fox, and R. Garnett, eds., *Advances in Neural Information Processing Systems 32*, pp. 8024–8035. Curran Associates, Inc., 2019.

[51] C. Rudin. Stop explaining black box machine learning models for high stakes decisions and use interpretable models instead. *Nature Machine Intelligence*, 1(5):206–215, 2019.

[52] C. Ruiz, M. Zitnik, and J. Leskovec. Identification of disease treatment mechanisms through the multiscale interactome. *Nature Communications*, 12(1):1–15, 2021.

[53] A. A. Sastre, R. W. Vernooij, M. G.-C. Harmand, and G. Martínez. Effect of the treatment of type 2 diabetes mellitus on the development of cognitive impairment and dementia. *Cochrane Database of Systematic Reviews*, (6), 2017.

[54] T. Schäfer, N. Kriege, L. Humbeck, K. Klein, O. Koch, and P. Mutzel. Scaffold hunter: a comprehensive visual analytics framework for drug discovery. *Journal of Cheminformatics*, 9(1):1–18, 2017.

[55] M. S. Schlichtkrull, N. De Cao, and I. Titov. Interpreting graph neural networks for nlp with differentiable edge masking. In *International Conference on Learning Representations*, 2020.

[56] T. Schnake, O. Eberle, J. Lederer, S. Nakajima, K. T. Schütt, K.-R. Müller, and G. Montavon. Higher-order explanations of graph neural networks via relevant walks. *arXiv preprint arXiv:2006.03589*, 2020.

[57] M. Sedlmair, M. Meyer, and T. Munzner. Design study methodology: Reflections from the trenches and the stacks. *IEEE Transactions on Visualization and Computer Graphics*, 18(12):2431–2440, 2012.

[58] A. Simkute, E. Luger, B. Jones, M. Evans, and R. Jones. Explainability for experts: A design framework for making algorithms supporting expert decisions more explainable. *Journal of Responsible Technology*, 7:100017, 2021.

[59] K. Sokol and P. Flach. Explainability fact sheets: a framework for systematic assessment of explainable approaches. In *Proceedings of the 2020 Conference on Fairness, Accountability, and Transparency*, pp. 56–67, 2020.

[60] D. N. Sosa, A. Derry, M. Guo, E. Wei, C. Brinton, and R. B. Altman. A literature-based knowledge graph embedding method for identifying drug repurposing opportunities in rare diseases. In *PACIFIC SYMPOSIUM ON BIOCOMPUTING 2020*, pp. 463–474. World Scientific, 2019.

[61] C. D. Spinner, R. L. Gottlieb, G. J. Criner, J. R. A. López, A. M. Cattelan, A. S. Viladomiu, O. Ogbuagu, P. Malhotra, K. M. Mullane, A. Castagna, et al. Effect of remdesivir vs standard care on clinical status at 11 days in patients with moderate COVID-19: a randomized clinical trial. *Journal of American Medical Association*, 324(11):1048–1057, 2020.

[62] A. D. Team. Ant design. `https://github.com/ant-design/ant-design/`.

[63] S. Tonekaboni, S. Joshi, M. D. McCradden, and A. Goldenberg. What clinicians want: contextualizing explainable machine learning for clinical end use. In *Machine Learning for Healthcare Conference*, pp. 359–380. PMLR, 2019.

[64] L. Van der Maaten and G. Hinton. Visualizing data using t-sne. *Journal of Machine Learning Research*, 9(11), 2008.

[65] J. van Wijk. The value of visualization. In *VIS 05. IEEE Visualization, 2005.*, pp. 79–86, 2005.

[66] P. Veličković, G. Cucurull, A. Casanova, A. Romero, P. Liò, and Y. Bengio. Graph attention networks. In *International Conference on Learning Representations*, 2018.

[67] G. Vilone and L. Longo. Explainable artificial intelligence: a systematic review. *arXiv preprint arXiv:2006.00093*, 2020.

[68] D. Wang, Q. Yang, A. Abdul, and B. Y. Lim. Designing theory-driven user-centric explainable ai. In *Proceedings of the 2019 CHI conference on human factors in computing systems*, pp. 1–15, 2019.

[69] Q. Wang, S. L'Yi, and N. Gehlenborg. Improving the utility and usability of visualization in ai-driven scientific discovery. 2022.

[70] Q. Wang, T. Mazor, T. A. Harbig, E. Cerami, and N. Gehlenborg. Threadstates: State-based visual analysis of disease progression. In *Proceedings of the IEEE VIS*. IEEE, 2021.

[71] Q. Wang, Y. Ming, Z. Jin, Q. Shen, D. Liu, M. J. Smith, K. Veeramachaneni, and H. Qu. Atmseer: Increasing transparency and controllability in automated machine learning. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*, pp. 1–12, 2019.

[72] Q. Wang, Z. Xu, Z. Chen, Y. Wang, S. Liu, and H. Qu. Visual analysis of discrimination in machine learning. *IEEE Transactions on Visualization and Computer Graphics*, 27(2):1470–1480, 2020.

[73] Q. Wang, J. Yuan, S. Chen, H. Su, H. Qu, and S. Liu. Visual genealogy of deep neural networks. *IEEE Transactions on Visualization and Computer Graphics*, 26(11):3340–3352, 2019.

[74] X. Wang, H. Ji, C. Shi, B. Wang, Y. Ye, P. Cui, and P. S. Yu. Heterogeneous graph attention network. In *The World Wide Web Conference*, pp. 2022–2032, 2019.

[75] X. Wang and M. Yin. *Are Explanations Helpful? A Comparative Study of the Effects of Explanations in AI-Assisted Decision-Making*, p. 318–328. Association for Computing Machinery, New York, NY, USA, 2021.

[76] S. Wehrend and C. Lewis. A problem-oriented classification of visualization techniques. In *Proceedings of the First IEEE Conference on Visualization: Visualization90*, pp. 139–143. IEEE, 1990.

[77] D. S. Wishart, Y. D. Feunang, A. C. Guo, E. J. Lo, A. Marcu, J. R. Grant, T. Sajed, D. Johnson, C. Li, Z. Sayeeda, et al. Drugbank 5.0: a major update to the drugbank database for 2018. *Nucleic acids research*, 46(D1):D1074–D1082, 2018.

[78] W. Yang, Z. Li, M. Liu, Y. Lu, K. Cao, R. Maciejewski, and S. Liu. Diagnosing concept drift with visual analytics. In *Proceedings of IEEE Conference on Visual Analytics Science and Technology*, pp. 12–23, 2020.

[79] R. Ying, D. Bourgeois, J. You, M. Zitnik, and J. Leskovec. GNNExplainer: generating explanations for graph neural networks. *Advances in Neural Information Processing Systems*, 32:9240, 2019.

[80] H. Yuan, H. Yu, S. Gui, and S. Ji. Explainability in graph neural networks: A taxonomic survey. *arXiv preprint arXiv:2012.15445*, 2020.

[81] H. Yuan, H. Yu, J. Wang, K. Li, and S. Ji. On explainability of graph neural networks via subgraph explorations. *arXiv preprint arXiv:2102.05152*, 2021.

[82] J. Yuan, C. Chen, W. Yang, M. Liu, J. Xia, and S. Liu. A survey of visual analytics techniques for machine learning. *Computational Visual Media*, pp. 1–34, 2020.

[83] S. Yun, M. Jeong, R. Kim, J. Kang, and H. J. Kim. Graph transformer networks. *Advances in Neural Information Processing Systems*, 32:11983–11993, 2019.

[84] X. Zeng, X. Song, T. Ma, X. Pan, Y. Zhou, Y. Hou, Z. Zhang, K. Li, G. Karypis, and F. Cheng. Repurpose open data to discover therapeutics for covid-19 using deep learning. *Journal of proteome research*, 19(11):4624–4636, 2020.

[85] M. Zhang and Y. Chen. Link prediction based on graph neural networks. *Advances in Neural Information Processing Systems*, 31:5165–5175, 2018.

[86] J. R. Zilke, E. Loza Mencía, and F. Janssen. Deepred–rule extraction from deep neural networks. In *International conference on discovery science*, pp. 457–473. Springer, 2016.

[87] A. Zytek, D. Liu, R. Vaithianathan, and K. Veeramachaneni. Sibyl: Understanding and addressing the usability challenges of machine learning in high-stakes decision making. *IEEE Transactions on Visualization and Computer Graphics*, 2021.