Transactions on Machine Intelligence 9(1) (2026) 1–15

ISSN Online: 2821-1693

Transactions on Machine Intelligence

Journal Homepage: https://tmachineintelligence.ir/index.php/TMI

MultiCNKG: Integrating Cognitive Neuroscience, Gene, and Disease Knowledge Graphs Using Large Language Models

A. Sarabadani [1,*], Kh. Rahsepar Fard [2], H. Hariri[3]

[1] Department of Computer Engineering and Information Technology, University of Qom, Qom, Iran.
[2] Department of Computer Engineering and Information Technology, University of Qom, Qom, Iran.
[3] Department of Electrical and Computer Engineering, Malek Ashtar University of Technology, Tehran, Iran.

ARTICLE INFO

Article History:
Received 26 November 2025
Received in revised form 4 January 2026
Accepted 22 February 2026
Available online 1 March 2026

Keywords:
Knowledge Graph, Large Language Models, Cognitive Neuroscience, Gene Ontology, Disease Ontology, Knowledge Integration

ABSTRACT

The advent of large language models (LLMs) has revolutionized the integration of knowledge graphs (KGs) in the biomedical and cognitive sciences, effectively overcoming the limitations of traditional machine learning methods in capturing intricate semantic links among genes, diseases, and cognitive processes. This paper introduces MultiCNKG, an innovative framework that merges three distinct knowledge sources: the Cognitive Neuroscience Knowledge Graph (CNKG), containing 2.9K nodes and 4.3K edges across 9 node types and 20 edge types; the Gene Ontology (GO), featuring 43K nodes and 75K edges across 3 node types and 4 edge types; and the Disease Ontology (DO), comprising 11.2K nodes and 8.8K edges with 1 node type and 2 edge types. Utilizing advanced LLMs such as GPT-4, we perform automated entity alignment, semantic similarity computation, and graph augmentation to construct a unified, cohesive KG that interconnects genetic mechanisms, neurological disorders, and cognitive functions. The resulting MultiCNKG unified graph encompasses 6.9K nodes across 5 distinct types and 11.3K edges spanning 7 relational types, establishing a multi-layered analytical pipeline from molecular to behavioral domains. Empirical evaluations demonstrate robust framework performance, achieving an 85.20% precision rate, 87.30% recall, 92.18% coverage, 82.50% graph consistency, a 40.28% novelty detection rate, and an 89.50% expert validation score. Furthermore, link prediction benchmarks utilizing TransE (MR: 391, MRR: 0.411) and RotatE (MR: 263, MRR: 0.395) yield highly competitive performance against standard benchmarks like FB15k-237 and WN18RR. Ultimately, this integrated KG advances clinical and research applications in personalized medicine, cognitive disorder diagnostics, and data-driven hypothesis formulation within cognitive neuroscience.

1. INTRODUCTION

Knowledge Graphs (KGs) have become foundational assets across numerous computational paradigms, effectively modeling complex relational entities within diverse real-world scenarios [1]. A significant subset of these structures such as scholarly citation networks [2] and e-commerce product graphs [3] manifest as Text-Attributed

* Corresponding author: a.sarabadani@stu.qom.ac.ir
Department of Computer Engineering and Information Technology, University of Qom, Qom, Iran.

http://dx.doi.org/10.22034/TMI.2026.244943

Graphs (TAGs), where structural nodes are intrinsically linked to rich textual metadata. For example, in benchmark datasets like ogbn-products, each discrete node represents a commercial product, utilizing its textual description as a high-dimensional node attribute [2]. These TAG architectures are widely deployed across multiple domains, including social network analysis [4], semantic information retrieval [5], and a broad spectrum of natural language processing (NLP) tasks. Among these, biomedical KGs are particularly vital due to their exceptional capacity to map and uncover the highly intricate, multi-layered interconnections governing biological systems, pathological diseases, and genetic mechanisms.

Driven by the prevalence and utility of TAGs, the primary objective of this investigation is to engineer a unified knowledge graph architecture that synthesizes three heterogeneous biomedical knowledge sources: the Cognitive Neuroscience Knowledge Graph (CNKG), the Disease Ontology (DO), and the Gene Ontology (GO).

Concurrently, Graph Neural Networks (GNNs) [6] have emerged as the standard algorithmic framework for processing graph-structured data, utilizing iterative message-passing mechanisms to capture local neighborhoods and topological structures. However, traditional GNN pipelines heavily rely on shallow, non-contextual embedding paradigms such as Bag-of-Words (BoW) [7] and Word2Vec [8] to encode textual node attributes. Recent literature underscores that these shallow representations are structurally restricted; they struggle to resolve polysemous lexical tokens [9] and lack the semantic depth required to capture complex context [10], which frequently degrades performance in downstream analytical tasks.

In contrast, Large Language Models (LLMs) offer extensive context-aware parametric knowledge and superior semantic comprehension, achieved via self-supervised pre-training on massive text corpora [8]. This pre-trained foundation has driven a major shift in downstream NLP tasks [11]. State-of-the-art architectures such as ChatGPT and GPT-4 [12], which leverage hundreds of billions of parameters, consistently demonstrate exceptional performance across diverse text-driven applications. This architectural evolution introduces a critical research question: Can the contextual and semantic capabilities of LLMs compensate for the structural and representative shortcomings of traditional GNN pipelines? Recent LLM-driven methodologies have increasingly replaced conventional text encoders, successfully leveraging pre-trained weights to excel in tasks characterized by implicit graph structures including personalized recommendation, semantic ranking, and multi-hop reasoning. This success highlights the capacity of LLMs to navigate complex, multi-modal relational challenges.

The rapid evolution of Transformer-based language models has fundamentally transformed the landscape of NLP and relational data mining. These deep architectures are structurally classified into encoder-only variants (e.g., BERT [13]), decoder-only networks (e.g., the GPT series [14]), and unified encoder-decoder topologies (e.g., T5 [15]). Frontier models including OpenAI's GPT-4 [12], Google's Gemini [16], PaLM [17], Microsoft's Phi-3 [18], and Meta's LLaMA [19] are collectively categorized as LLMs. Engineered on large-scale Transformer blocks with billions of parameters, these networks manifest advanced agentic capabilities, spanning automated perception, multi-step logical reasoning, strategic planning, and autonomous action execution [20].

Capitalizing on these paradigms, the primary contribution of this study is the development of MultiCNKG-LLM, an innovative framework designed to systematically integrate data from three distinct knowledge bases: CNKG, DO, and GO. The MultiCNKG-LLM framework not only preserves the critical, explicit relationships inherent in the source datasets but also leverages the semantic reasoning of LLMs to uncover novel, previously hidden connections. This integrated approach represents a substantial advancement in the scalability and utility of biomedical knowledge graphs.

2. RELATED WORKS

Knowledge Graphs (KGs) have emerged as powerful frameworks for representing, structuring, and integrating highly heterogeneous biomedical data. By mapping complex biological spaces, these architectures facilitate efficient data synthesis, allowing researchers to discover novel, non-trivial associations among distinct entities such as pathological diseases, genetic sequences, and therapeutic drugs. Notable pioneer benchmarks in this domain include open-source platforms like Bio2RDF [21] and specialized frameworks such as the Clinical Trial Knowledge Graph (CTKG) developed by Chen et al. [22].

Concurrently, the paradigm of biomedical knowledge processing has been fundamentally reshaped by the deployment of Large Language Models (LLMs) optimized for clinical text. These domain-specific models have drastically advanced the structural interpretation of dense medical literature, unstructured clinical narratives, and diagnostic decision-support datasets. Prominent frontier architectures in healthcare include DeepMind's MedIC, Microsoft's BioGPT [23], TrialGPT [24], BioBART [25], and BioMistral [26]. By efficiently processing massive volumes of unstructured medical data, these models have established a profound impact across drug discovery pipelines, patient-specific EHR data analytics, and real-time clinical decision-making.

Recent investigations underscore the efficacy of combining generative models with topological structures to construct advanced systems such as DIAMOND-KG a knowledge graph engineered to capture specific drug indications alongside their corresponding clinical context [27]. Empirical evidence from that study demonstrates that a sophisticated, LLM-driven text extraction strategy substantially enhances the overall structural quality of the KG; notably, 71% of extracted drug indications were successfully linked to at least one relevant clinical reference, emphasizing the critical value of source-grounded reference data in elevating KG utility and trust metrics.

The convergence of LLMs and KGs is expanding rapidly across medicine, neuroscience, and advanced biomedicine, establishing new frontiers for automated knowledge acquisition. Researchers have developed complex platforms, such as the framework introduced by Mann et al. [28], which systematically aggregate disparate, legacy medical knowledge repositories into a singular, unified KG to support downstream treatment identification based on co-occurring symptoms or disease attributes.

To accelerate text processing, fast LLM inference techniques have gained significant traction in automated clinical concept extraction [29], noticeably raising the recall rate of entity extraction from unstructured EHR data. Nonetheless, core challenges persist particularly regarding the model's capacity to distinguish ambiguous alphanumeric drug codes within fine-grained medical contexts. Despite recent algorithmic progress, many automated pipelines still struggle to accurately resolve these critical details.

Beyond standalone extraction, recent studies highlight the significant potential of LLMs to execute graph-based tasks directly, including relationship extraction, KG completion, and semantic graph querying [30]. For instance, Khorashadizadeh et al. [31] conducted a rigorous qualitative analysis utilizing decoder-only models, such as the ChatGPT and Bard series, to evaluate KG structural integrity and execute complex biomedical queries. Their findings indicated that while LLMs show exceptional promise in automated KG extraction pipelines, the associated computational overhead and inference costs remain challenges for scalable deployment.

With the rapid ascent of generative AI, the automated evaluation of KG health models has evolved into a key research domain. Contemporary studies indicate that fine-tuned LLMs can automatically assess and validate the structural consistency of KG schemas [32], presenting a scalable, automated alternative to traditional, labor-intensive human verification. Furthermore, LLMs have been successfully deployed to audit class-member relationships within massive public KGs like Wikidata and Calligraph, providing knowledge engineers with automated diagnostics to enhance taxonomic precision [33].

In parallel, various semantic optimization techniques have been engineered to streamline KG operations through embedded vector similarity searches. Modern architectures frequently implement dense embedded searches followed by localized structural retrieval, or leverage query classification layers to organize graph filters and summarize LLM outputs [34-35]. More advanced neuro-symbolic approaches incorporate structural neighborhood descriptors, automated query rewriting, logical query decomposition, and the direct fine-tuning of model weights using graph-structured training pairs [36-39].

Finally, LLMs have demonstrated remarkable empirical performance gains in direct clinical tasks, such as automated disease diagnosis and prognostic prediction utilizing electronic medical records (EMRs). For example, Jiang et al. [40] developed a patient-specific KG leveraging a bidirectional attention-enhanced graph neural network (BAT-GNN) to improve clinical decision-support systems. Similarly, the RAM-EHR framework converts heterogeneous clinical knowledge sources into structured text formats via multi-modal co-training, optimizing information extraction to yield highly accurate predictive medical metrics [41].

To improve diagnostic interpretability, systems like DR. KNOWS integrate dense KGs with formal clinical logic models grounded in the Unified Medical Language System (UMLS), successfully bridging the gap between symbolic reasoning and neural representations [42]. Furthermore, the REALM architecture addresses complex, multi-faceted medical cases by seamlessly integrating longitudinal patient records and multivariate time-series data using Retrieval-Augmented Generation (RAG) technology, ensuring computationally efficient and context-grounded medical insights.

3. METHODOLOGY

The proposed methodology for MultiCNKG consists of four major stages:

1. Data Collection and Preprocessing
2. Knowledge Graph Representation
3. Graph Alignment and Integration
4. Graph Expansion using Large Language Models (LLMs)

Each stage is described in detail below.

3.1. Data Collection and Preprocessing

Three knowledge sources were selected:

- Cognitive Neuroscience Knowledge Graph (CNKG): Represents concepts, tasks, and processes in cognitive neuroscience, such as memory, attention, decision-making, and learning, along with their neural correlates.
- Gene Ontology (GO): Provides structured annotations of genetic and molecular functions.
- Disease Ontology (DO): Encodes diseases and disorders, including neurological and psychiatric conditions.

Table 1.CNKG , DO and GO

Name	# Node	# Node Types	# Edges	# Edge Types
Cognitive Neuroscience Knowledge Graph (CNKG)	2.9 K	9	4.3 K	20
Disease Ontology	11.2 K	1	8.8 K	2
Gene Ontology	43 K	3	75 K	4

The datasets were preprocessed using LLM-based pipelines (GPT-4, BioGPT). Preprocessing steps included:

Tokenization: Raw text descriptions of entities were tokenized into sequences:

$$T = \{t_1. t_2. \ldots . t_n\} \quad (1)$$

where each ti represents a token.

Normalization: To ensure consistency, entities with multiple forms (e.g., Alzheimer's disease and AD) were mapped to canonical forms using embedding similarity:

$$\text{sim}(e_i. e_j) = \frac{v_i \cdot v_j}{\|v_i\|\|v_j\|} \quad (2)$$

where vi, vj are vector embeddings generated by LLMs.

Noise Removal: Duplicate or irrelevant nodes were removed using threshold-based filtering on similarity scores.

3.2. Knowledge Graph Representation

The integrated KG is defined as a directed multigraph:

$$G = (V.E.R) \tag{3}$$

where:

V is the set of nodes (genes, diseases, cognitive neuroscience concepts),

$E \subseteq V \times R \times V$ is the set of edges,

R is the set of relations (e.g., causes, associated_with, impairs).

Each edge is represented as a knowledge triple:

$$(h.r.t)\ with\ h.t \in V.r \in R \tag{4}$$

For example:

(APOE4, causes, Alzheimer's disease)

(Alzheimer's disease, impairs, Memory)

The graph structure can also be encoded in an adjacency matrix:

$$A_{ij} = \begin{cases} 1 & \textit{if there exists a relation between } v_i \textit{ and } v_j \\ 0 & \textit{otherwise} \end{cases} \tag{5}$$

3.3. Graph Alignment and Integration

To merge CNKG, GO, and DO, we align nodes and edges that represent the same or semantically related entities.

1. Node Alignment: For nodes $vi \in VCNKG$, $vj \in VDO$, and $vk \in VGO$, the final unified set is:

   $$V_f = V_{CNKG} \cup V_{DO} \cup V_{GO} \tag{6}$$

   Similar nodes are merged if their embedding similarity satisfies:

   $$\mathrm{sim}(\mathrm{v_i}.\mathrm{v_j}) \geq \tau \tag{7}$$

   where τ\tau is a predefined threshold.

2. Edge Alignment: Edges that represent equivalent relations (e.g., causes vs. induces) are unified by LLM-based semantic comparison. The final edge set is:

   $$E_f = E_{CNKG} \cup E_{DO} \cup E_{GO} \tag{8}$$

3. Adjacency Update: The adjacency matrix is updated as:

   $$A_f = A_{CNKG} + A_{DO} + A_{GO} + \Delta A \tag{9}$$

where ΔA represents newly discovered relations.

3.4. Graph Expansion with LLMs

The most critical step is leveraging LLMs to discover new relationships that are not explicitly present in the original graphs.

Relation Prediction: For two nodes $h, t \in V$, the probability of a new relation $r\ new$ is estimated by LLM:

$$P(r_{new} \mid h.t) = f_{LLM}(h.t) \tag{10}$$

A relation is added if:

$$P(r_{new} \mid h.t) \geq \tau \tag{11}$$

Probabilistic Similarity Model: Alternatively, similarity-based relation discovery is modeled as:

$$P(r_{new}) = 1 - exp\left(-\frac{\|v_h - v_t\|^2}{2\sigma^2}\right) \tag{12}$$

where vh , vt are embeddings of nodes $h, t, and\ \sigma$ controls distribution spread.

Iterative Expansion: The KG evolves iteratively:

$$G^{(t+1)} = G^{(t)} + \Delta G^{(t)} \tag{13}$$

Where $\Delta G^{(t)} = \left(\Delta V^{(t)}.\Delta E^{(t)}\right)$ are new entities and edges discovered at iteration tt.

Validation: Newly discovered relations are assigned a confidence score and validated either automatically via external databases or manually.

by domain experts. If feedback decreases confidence below threshold:

$$P(r_{new}) \leftarrow P(r_{new}) - \delta P \tag{14}$$

the relation is removed in the next iteration.

3.5. Final Integrated Graph

The final MultiCNKG integrates three layers:

1. Genetic Layer (GO): Molecular and functional gene information.
2. Disease Layer (DO): Neurological and psychiatric disorders.
3. Cognitive Neuroscience Layer (CNKG): Memory, attention, learning, and decision-making constructs.

The integration allows tracing pathways from genes → diseases → cognitive functions, offering a holistic perspective for cognitive neuroscience research.

4. SIMULATION AND RESULTS

LLM was used to build the knowledge graph and both types of nodes and edges for MultiCNKG. This linguistic model extracted entities and edges from raw texts related to cognitive neuroscience, gene ontology, and disease ontology articles. This model identified 5 different node types (refer to Table 2) and 7 different edge types (refer to

Table 3). The strength of this model lies in its training to recognize diverse biomedical and cognitive neuroscience terms and their contexts, allowing for the distinction between various types of nodes and relationships.

Table 2. Five Different Node Types Extracted by LLM from MultiCNKG-Related Texts.

#	Node Types	Use
1	Genes	Represents genetic factors linked to cognitive functions and diseases (e.g., APOE, BDNF).
2	Diseases	Represents neurological and cognitive disorders (e.g., Alzheimer's, Parkinson's).
3	Cognitive Processes	Represents higher-order cognitive functions (e.g., memory, attention).
4	Biological Pathways	Represents molecular pathways involved in cognition and disease (e.g., synaptic plasticity).
5	Therapeutic Targets	Represents potential targets for intervention (e.g., neurotransmitter receptors).

The evaluation of MultiCNKG was conducted using various metrics derived from the integrated knowledge graph. MultiCNKG, with 6.9K nodes and 11.3K edges, demonstrates a balanced representation with 5 node types and 7 edge types, reflecting its focus on integrating cognitive neuroscience, genetic, and disease data. The model leverages LLM capabilities to uncover new relationships, showcasing its semantic coherence and novelty. However, challenges such as scalability and the reliance on proprietary LLMs like GPT-4 highlight areas for improvement. Future enhancements could involve adopting open-source models to enhance transparency and address computational constraints, ensuring robust KG construction for cognitive neuroscience applications.

Table 3. Five Different Edge Types Extracted by LLM from MultiCNKG-Related Texts.

#	**Edge Types**	**Use**
1	Causes	Indicates causation between entities (e.g., gene mutation causes disease).
2	Associated with	General associations between entities (e.g., gene associated with cognitive decline).
3	Regulates	Indicates regulatory relationships (e.g., gene regulates pathway).
4	Involved in	Represents involvement in processes or diseases (e.g., pathway involved in memory).
5	Treated by	Indicates treatment relationships (e.g., target treated by drug).
6	Influences	Represents influence on cognitive or biological entities (e.g., gene influences cognition).
7	Linked to	Connects entities across domains (e.g., disease linked to cognitive process).

5. EVALUATION

The performance of the MultiCNKG knowledge graph is evaluated in two ways. In the first method, we evaluate using traditional evaluation criteria. In the second method, the evaluation of LLM includes using large language models such as GPT-4. The model evaluation process is shown in Table 4.

Table 4. Details related to various knowledge graphs and knowledge graphs of the proposed approach.

Name	# Node	# Node Types	# Edges	# Edge Types
DrKG [43]	97 K	13	5.8 M	107
PrimeKG [44]	129.4 K	10	8.1 M	30
Gene Ontology [45]	43 K	3	75 K	4
GP-KG [46]	61.1 K	7	124 K	9
DDKG [47]	551	2	2.7 K	1
Disease Ontology[48]	11.2 K	1	8.8 K	2
DrugBank [49]	7.4 K	4	366 K	4
PharmKG [50]	7.6 K	3	500 K	3
MultiCNKG	6.9 K	5	11.3 K	7

Our evaluation of MultiCNKG was based on criteria such as precision, recall, F1 score, coverage, graph consistency, computational efficiency, novelty detection, and expert validation. In the second part, we evaluated the ability of the model to predict links, using criteria such as Mean Rank (MR), Mean Reciprocal Rank (MRR), and Precision at K (P@K).

5.1. Traditional Evaluation Methods

Accuracy, precision, recall and F1-score measure the quality of the node and edge alignment processes. The higher the precision, the fewer incorrect alignments (false positives); the higher the recall, the fewer missed alignments (false negatives).

$$Precision = \frac{TP}{TP + FP} \tag{15}$$

$$Recall = \frac{TP}{TP + FN} \tag{16}$$

$$F1\ Score = 2 \times \frac{recall \times precision}{recall + precision} \tag{17}$$

In addition, the following metrics were also used to evaluate the proposed model:

Coverage: proportion of original KGs (CNKG, GO, DO) that have been successfully transferred to the final merged KG.

$$Coverage = \frac{Unique_{Nodes}(N)_{merged} + Unique_{Edges}(E)_{merged}}{\sum_{i=1}^{n} (N_i + E_i)} \tag{18}$$

Graph Consistency: evaluated through semantic and structural checks (ontology-based similarity and OWL-based consistency).

Computational Efficiency: measured in terms of time and space complexity during the integration process.

Novelty Detection: evaluates the ability of the integrated KG to discover new and previously unknown relationships.

$$Novelty_{Score} = \frac{NewlyDiscovered_{Edges}(E)_{LLM}}{Total_{Edges}(E)_{merged}} \tag{19}$$

5.2. Expert Validation

Expert Validation: domain experts verified whether new relations suggested by LLMs were biologically and cognitively plausible.Experts in cognitive neuroscience and biomedicine played a critical role in validating the biological and clinical relevance of new proposed edges. Validation results confirmed that most suggested relations were meaningful and scientifically plausible.

For example:

- GO vs. MultiCNKG: Precision slightly decreased, but Recall and Coverage improved.
- DO vs. MultiCNKG: Results were close, but MultiCNKG had better novelty and expert validation.
- CNKG vs. MultiCNKG: MultiCNKG showed better integration by linking cognitive constructs with genetic and disease information.

Table 5 illustrate comparisons between individual KGs (GO, DO, CNKG) and the integrated MultiCNKG approach.

Table 5. Summarizes the evaluation results.

Metrics	Gene Ontology (GO)	Disease Ontology (DO)	CNKG	MultiCNKG
Precision	89.54	83.74	80.12	85.20
Recall	79.48	85.24	82.60	87.30
F1-Score	83.31	84.48	81.35	80.15
Coverage	89.52	73.09	78.41	92.18
Graph Consistency	75.24	71.69	79.11	82.50
Computational Eff.	53.01	70.17	62.42	68.90
Novelty Detection	44.61	41.05	46.38	40.28
Expert Validation	84.78	90.22	88.71	89.50

5.3. Link Prediction

Different evaluation dimensions were covered by various metrics to comprehensively evaluate MultiCNKG. Link prediction methods such as TransE, RotatE, DistMult, ComplEx, ConvE, and HolmE were applied to assess predictive strength.

The following metrics were used [51]:

Mean Rank (MR):

$$MR = \frac{1}{N}\sum_{i=1}^{N} rank_i \tag{20}$$

Mean Reciprocal Rank (MRR):

$$MRR = \frac{1}{N}\sum_{i=1}^{N} \frac{1}{rank_i} \tag{21}$$

Precision at K (P@K):

$$P@K = \frac{1}{N}\sum_{i=1}^{N} \frac{Number\ of\ relevant\ items\ in\ top\ K}{K} \tag{22}$$

Experimental results showed that MultiCNKG achieves comparable or superior results to individual graphs (GO, DO, CNKG), particularly in terms of novelty detection and expert validation.

Table 6. summarizes the evaluation Link Prediction.

KG		TransE [55]	RotatE [56]	DistMult [57]	ComplEx [58]	ConvE [59]	HolmE[6[60]
FB15k-237 [52]	MR	209	178	199	144	281	-
	MRR	0.310	0.336	0.313	0.367	0.305	0.331
	P@1	0.217	0.238	0.224	0.271	0.219	0.237
	P@3	0.257	0.328	0.263	0.275	0.350	0.366
	P@10	0.496	0.530	0.490	0.558	0.476	0.517
WN18RR [53]	MR	3936	3318	5913	2867	4944	-
	MRR	0.206	0.475	0.433	0.489	0.427	0.466
	P@1	0.279	0.426	0.396	0.442	0.389	0.415
	P@3	0.364	0.492	0.440	0.460	0.430	0.489
	P@10	0.495	0.573	0.502	0.580	0.507	0.561
YAGO3-10 [54]	MR	1187	1830	1107	793	2429	-
	MRR	0.501	0.498	0.501	0.577	0.488	0.441
	P@1	0.405	0.405	0.412	0.500	0.399	0.333
	P@3	0.528	0.550	0.38	0.40	0.560	0.507
	P@10	0.673	0.670	0.661	0.7129	0.657	0.641
MultiCNKG	MR	391	263	242	212	375	-
	MRR	0.411	0.395	0.418	0.301	0.321	0.211
	P@1	0.435	0.342	0.444	0.333	0.329	0.301
	P@3	0.499	0.463	0.475	0.349	0.363	0.336
	P@10	0.518	0.571	0.519	0.487	0.493	0.403

Table 6 summarizes the link prediction evaluation results across multiple knowledge graphs and embedding models. The results show that COMPLEX and ROTATE generally achieve the highest Mean Reciprocal Rank (MRR) and precision metrics, particularly on FB15K-237 and WN18RR datasets, indicating their effectiveness in capturing complex relational patterns. For YAGO3-10, COMPLEX again demonstrates superior performance with an MRR of 0.577, reflecting its robustness in handling multi-relational data. In the proposed MULTICNKG, which integrates concepts from recent LLM-based knowledge graph frameworks, performance remains competitive, especially in P@1 and P@10 metrics under TRANSE and DISTMULT, showing balanced precision and recall characteristics. These findings align with recent advances in LLM-assisted graph construction and reasoning. Specifically, the DKG-LLM framework [61] employs dynamic knowledge integration for medical diagnosis and treatment personalization, the ExKG-LLM model [62] enhances automated expansion of cognitive neuroscience knowledge graphs with improved precision and coverage, and the GDPKG-LLM framework [63] effectively unifies gene, disease, and pharmacogenomic knowledge using GPT-4. Together, these studies demonstrate the growing synergy between LLMs and graph-based models, where advanced embedding and reasoning mechanisms contribute to improved link prediction accuracy and richer semantic representation across biomedical and cognitive domains.

6. CONCLUSION AND FUTURE WORK

In this study, we have successfully developed MultiCNKG, a unified knowledge graph that integrates cognitive neuroscience concepts from CNKG, genetic annotations from GO, and disease representations from DO, resulting in a comprehensive resource with 6.9K nodes across 5 distinct types (Genes, Diseases, Cognitive Processes, Biological Pathways, and Therapeutic Targets) and 11.3K edges spanning 7 relation types (Causes, Associated with, Regulates, Involved in, Treated by, Influences, and Linked to). By employing large language models (LLMs) such as GPT-4 for entity alignment, semantic analysis, and graph expansion, MultiCNKG not only preserves essential relationships from the source graphs but also uncovers novel connections, bridging molecular genetics with neurological diseases and higher-order cognitive functions. Comparative analysis with other biomedical KGs (e.g., DrKG with 97K nodes and 5.8M edges, PrimeKG with 129.4K nodes and 8.1M edges) highlights MultiCNKG's focused efficiency, while evaluation metrics demonstrate superior performance in recall (87.30%), coverage (92.18%), graph consistency (82.50%), and expert validation (89.50%) compared to individual sources like GO (recall: 79.48%, coverage: 89.52%) and DO (recall: 85.24%, coverage: 73.09%). Link prediction results further validate its predictive power, with models like ComplEx achieving a Mean Rank (MR) of 212 and Mean Reciprocal

Rank (MRR) of 0.301, outperforming or matching benchmarks on datasets such as FB15k-237 (ComplEx MR: 144, MRR: 0.367) and WN18RR (ComplEx MR: 2867, MRR: 0.489) in targeted metrics like Precision@10 (0.487). These outcomes underscore MultiCNKG's semantic coherence, novelty in relationship discovery, and potential for enabling holistic insights in cognitive neuroscience, from gene-disease pathways to behavioral impacts.

Despite these advancements, challenges remain, including scalability for larger datasets and dependency on proprietary LLMs, which may limit accessibility and transparency. Future work will explore incorporating open-source LLMs (e.g., BioGPT or LLaMA variants) to reduce computational costs and enhance reproducibility. We plan to expand MultiCNKG by integrating additional ontologies, such as DrugBank (7.4K nodes, 366K edges) or PharmKG (7.6K nodes, 500K edges), to support drug repurposing and therapeutic predictions. Further enhancements could involve real-time graph updates using federated learning, advanced validation through crowdsourced expert feedback, and application-specific extensions for personalized medicine, such as AI-driven early detection of cognitive disorders like Alzheimer's. Ultimately, MultiCNKG sets a foundation for interdisciplinary research, fostering innovations in hypothesis generation and clinical decision-making in biomedicine and cognitive sciences.

Declaration

We acknowledge that we used ChatGPT to enhance the academic writing of our manuscript while ensuring the originality and integrity of our work.

Transparency Statement

The data supporting this study are available upon reasonable request to the corresponding author, subject to ethical and confidentiality considerations.

Acknowledgments

We would like to express our gratitude to all individuals who contributed to this project.

Declaration of Interest

The authors declare that they have no competing interests.

Funding

This research received no specific grant from any funding agency, commercial, or not-for-profit sectors.

REFERENCES

[1] Xia, F., Sun, K., Yu, S., Aziz, A., Wan, L., Pan, S., & Liu, H. (2021). Graph learning: A survey. IEEE Transactions on Artificial Intelligence, 2(2), 109-127. https://doi.org/10.1109/TAI.2021.3076021

[2] Hu, W., Fey, M., Zitnik, M., Dong, Y., Ren, H., Liu, B., Catasta, M., & Leskovec, J. (2020). Open graph benchmark: Datasets for machine learning on graphs. Advances in Neural Information Processing Systems, 33, 22118-22133.

[3] Chiang, W. L., Liu, X., Si, S., Li, Y., Bengio, S., & Hsieh, C. J. (2019). Cluster-GCN: An efficient algorithm for training deep and large graph convolutional networks. Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining, 257-266. https://doi.org/10.1145/3292500.3330925

[4] Li, Q., Li, X., Chen, L., & Wu, D. (2022). Distilling knowledge on text graph for social media attribute inference. Proceedings of the 45th International ACM SIGIR Conference on Research and Development in Information Retrieval, 2024-2028. https://doi.org/10.1145/3477495.3531968

[5] Zhu, J., Cui, Y., Liu, Y., Sun, H., Li, X., Pelger, M., Yang, T., Zhang, L., Zhang, R., & Zhao, H. (2021). TextGNN: Improving text encoder via graph neural network in sponsored search. Proceedings of the Web Conference 2021, 2848-2857. https://doi.org/10.1145/3442381.3449842

[6] Ma, Y., & Tang, J. (2021). Deep learning on graphs. Cambridge University Press. https://doi.org/10.1017/9781108924184

[7] Shao, Y., Taylor, S., Marshall, N., Morioka, C., & Zeng-Treitler, Q. (2018). Clinical text classification with word embedding features vs. bag-of-words features. 2018 IEEE International Conference on Big Data (Big Data), 2874-2878. IEEE. https://doi.org/10.1109/BigData.2018.8622345

[8] Mikolov, T. (2013). Efficient estimation of word representations in vector space. arXiv preprint arXiv:1301.3781.

[9] Qiu, X., Sun, T., Xu, Y., Shao, Y., Dai, N., & Huang, X. (2020). Pre-trained models for natural language processing: A survey. Science China Technological Sciences, 63(10), 1872-1897. https://doi.org/10.1007/s11431-020-1647-3

[10] Miaschi, A., & Dell'Orletta, F. (2020). Contextual and non-contextual word embeddings: An in-depth linguistic investigation. Proceedings of the 5th Workshop on Representation Learning for NLP, 110-119. https://doi.org/10.18653/v1/2020.repl4nlp-1.15

[11] Zhao, W. X., Zhou, K., Li, J., Tang, T., Wang, X., Hou, Y., Min, Y., Zhang, B., Zhang, J., Dong, Z., et al. (2023). A survey of large language models. arXiv preprint arXiv:2303.18223.

[12] Achiam, J., Adler, S., Agarwal, S., Ahmad, L., Akkaya, I., Aleman, F. L., et al. (2023). GPT-4 technical report. arXiv preprint arXiv:2303.08774.

[13] Kenton, J. D. M. W. C., & Toutanova, L. K. (2019). BERT: Pre-training of deep bidirectional transformers for language understanding. Proceedings of NAACL-HLT, 1, 2. Minneapolis, Minnesota.

[14] Radford, A., Wu, J., Child, R., Luan, D., Amodei, D., & Sutskever, I., et al. (2019). Language models are unsupervised multitask learners. OpenAI Blog, 1(8), 9.

[15] Raffel, C., et al. (2020). Exploring the limits of transfer learning with a unified text-to-text transformer. Journal of Machine Learning Research, 21, 1-67.

[16] Team G., Anil, R., Borgeaud, S., Wu, Y., Alayrac, J. B., Yu, J., Soricut, R., Schalkwyk, J., Dai, A. M., Hauth, A., et al. (2023). Gemini: A family of highly capable multimodal models. arXiv preprint arXiv:2312.11805.

[17] Chowdhery, A., Narang, S., Devlin, J., Bosma, M., Mishra, G., Roberts, A., Barham, P., Chung, H. W., Sutton, C., Gehrmann, S., et al. (2023). PaLM: Scaling language modeling with pathways. Journal of Machine Learning Research, 24(240), 1-113.

[18] Abdin, M., Jacobs, S. A., Awan, A. A., Aneja, J., Awadallah, A., Awadalla, H., et al. (2024). Phi-3 technical report: A highly capable language model locally on your phone. arXiv preprint arXiv:2404.14219.

[19] Touvron, H., Lavril, T., Izacard, G., Martinet, X., Lachaux, M. A., Lacroix, T., Rozière, B., Goyal, N., Hambro, E., Azhar, F., et al. (2023). LLaMA: Open and efficient foundation language models. arXiv preprint arXiv:2302.13971.

[20] Webb, T., Holyoak, K. J., & Lu, H. (2023). Emergent analogical reasoning in large language models. Nature Human Behaviour, 7(9), 1526-1541. https://doi.org/10.1038/s41562-023-01659-w

[21] Belleau, F., Nolin, M. A., Tourigny, N., Rigault, P., & Morissette, J. (2008). Bio2RDF: Towards a mashup to build bioinformatics knowledge systems. Journal of Biomedical Informatics, 41(5), 706-716. https://doi.org/10.1016/j.jbi.2008.03.004

[22] Himmelstein, D. S., Lizee, A., Hessler, C., Brueggeman, L., Chen, S. L., Hadley, D., Green, A., Khankhanian, P., & Baranzini, S. E. (2017). Systematic integration of biomedical knowledge prioritizes drugs for repurposing. eLife, 6, e26726. https://doi.org/10.7554/eLife.26726.017

[23] Luo, R., Sun, L., Xia, Y., Qin, T., Zhang, S., Poon, H., & Liu, T. Y. (2022). BioGPT: Generative pre-trained transformer for biomedical text generation and mining. Briefings in Bioinformatics, 23(6), bbac409. https://doi.org/10.1093/bib/bbac409

[24] Jin, Q., Wang, Z., Floudas, C. S., Chen, F., Gong, C., Bracken-Clarke, D., Xue, E., Yang, Y., Sun, J., & Lu, Z. (2023). Matching patients to clinical trials with large language models. arXiv. https://doi.org/10.1038/s41467-024-53081-z

[25] Yuan, H., Yuan, Z., Gan, R., Zhang, J., Xie, Y., & Yu, S. (2022). BioBART: Pretraining and evaluation of a biomedical generative language model. arXiv preprint arXiv:2204.03905. https://doi.org/10.18653/v1/2022.bionlp-1.9

[26] Labrak, Y., Bazoge, A., Morin, E., Gourraud, P. A., Rouvier, M., & Dufour, R. (2024). BioMistral: A collection of open-source pretrained large language models for medical domains. arXiv preprint arXiv:2402.10373. https://doi.org/10.18653/v1/2024.findings-acl.348

[27] Alharbi, R., Ahmed, U., Dobriy, D., Lajewska, W., Menotti, L., Saeedizade, M. J., & Dumontier, M. (2023). Exploring the role of generative AI in constructing knowledge graphs for drug indications with medical context. Proceedings http://ceur-ws.org.

[28] Wawrzik, F., Rafique, K. A., Rahman, F., & Grimm, C. (2023). Ontology learning applications of knowledge base construction for microelectronic systems information. Information, 14(3), 176. https://doi.org/10.3390/info14030176

[29] Peng, C., Yang, X., Yu, Z., Bian, J., Hogan, W. R., & Wu, Y. (2023). Clinical concept and relation extraction using prompt-based machine reading comprehension. Journal of the American Medical Informatics Association, 30(9), 1486-1493. https://doi.org/10.1093/jamia/ocad107

[30] Pan, S., Luo, L., Wang, Y., Chen, C., Wang, J., & Wu, X. (2024). Unifying large language models and knowledge graphs: A roadmap. IEEE Transactions on Knowledge and Data Engineering. https://doi.org/10.1109/TKDE.2024.3352100

[31] Khorashadizadeh, H., Mihindukulasooriya, N., Tiwari, S., Groppe, J., & Groppe, S. (2023). Exploring in-context learning capabilities of foundation models for generating knowledge graphs from text. arXiv preprint arXiv:2305.08804.

[32] Boylan, J., Mangla, S., Thorn, D., Ghalandari, D. G., Ghaffari, P., & Hokamp, C. (2024). KGValidator: A Framework for Automatic Validation of Knowledge Graph Construction. arXiv preprint arXiv:2404.15923.

[33] Allen, B. P., & Groth, P. T. (2024). Evaluating Class Membership Relations in Knowledge Graphs using Large Language Models. arXiv preprint arXiv:2404.17000. https://doi.org/10.1007/978-3-031-78952-6_2

[34] Soman, K., Rose, P. W., Morris, J. H., Akbas, R. E., Smith, B., Peetoom, B., Villouta-Reyes, C., Cerono, G., Shi, Y., Rizk-Jackson, A., et al. (2023). Biomedical knowledge graph-enhanced prompt generation for large language models. arXiv preprint arXiv:2311.17330. https://doi.org/10.1093/bioinformatics/btae560

[35] Cm, S., Prakash, J., & Singh, P. K. (2023). Question answering over knowledge graphs using BERT based relation mapping. Expert Systems, 40(10), e13456. https://doi.org/10.1111/exsy.13456

[36] Guo, Q., Cao, S., & Yi, Z. (2022). A medical question answering system using large language models and knowledge graphs. International Journal of Intelligent Systems, 37(11), 8548-8564. https://doi.org/10.1002/int.22955

[37] Wu, Y., Hu, N., Bi, S., Qi, G., Ren, J., Xie, A., & Song, W. (2023). Retrieve-rewrite-answer: A KG-to-text enhanced LLMs framework for knowledge graph question answering. arXiv preprint arXiv:2309.11206.

[38] Choudhary, N., & Reddy, C. K. (2023). Complex logical reasoning over knowledge graphs using large language models. arXiv preprint arXiv:2305.01157.

[39] Varshney, D., Zafar, A., Behera, N. K., & Ekbal, A. (2023). Knowledge grounded medical dialogue generation using augmented graphs. Scientific Reports, 13(1), 3310. https://doi.org/10.1038/s41598-023-29213-8

[40] Jiang, P., Xiao, C., Cross, A., & Sun, J. (2023). Graphcare: Enhancing healthcare predictions with personalized knowledge graphs. arXiv preprint arXiv:2305.12788.

[41] Xu, R., Shi, W., Yu, Y., Zhuang, Y., Jin, B., Wang, M. D., Ho, J. C., & Yang, C. (2024). RAM-EHR: Retrieval augmentation meets clinical predictions on electronic health records. arXiv preprint arXiv:2403.00815. https://doi.org/10.18653/v1/2024.acl-short.68

[42] Gao, Y., Li, R., Croxford, E., Tesch, S., To, D., Caskey, J., Patterson, B. W., Churpek, M. M., Miller, T., Dligach, D., et al. (2023). Large language models and medical knowledge grounding for diagnosis prediction. medRxiv, 2023-11. https://doi.org/10.1101/2023.11.24.23298641

[43] V. N. Ioannidis et al., "Few-shot Link Prediction via Graph Neural Networks for COVID-19 Drug-repurposing," 2020. DOI:10.48550/arXiv.2007.10261

[44] P. Chandak, K. Huang, and M. Zitnik, "Building a Knowledge Graph to Enable Precision Medicine," Scientific Data, 2023. https://doi.org/10.1038/s41597-023-01960-3

[45] M. Ashburner et al., "Gene Ontology: Tool for the Unification of Biology," Nature Genetics, vol. 25, no. 1, 2000, pp. 25-29. https://doi.org/10.1038/75556

[46] Z. Gao, P. Ding, and R. Xu, "KG-Predict: A Knowledge Graph Computational Framework for Drug Repurposing," Journal of Biomedical Informatics, vol. 132, 2022, 104133. https://doi.org/10.1016/j.jbi.2022.104133

[47] Z. Ghorbanali et al., "DrugRep-KG: Toward Learning a Unified Latent Space for Drug Repurposing Using Knowledge Graphs," Journal of Chemical Information and Modeling, vol. 63, no. 8, 2023, pp. 2532-2545. https://doi.org/10.1021/acs.jcim.2c01291

[48] L. Schriml et al., "Disease Ontology: A Backbone for Disease Semantic Integration," Nucleic Acids Research, vol. 40, no. D1, 2011, pp. D940-D946. https://doi.org/10.1093/nar/gkr972

[49] D. S. Wishart et al., "DrugBank 5.0: A Major Update to the DrugBank Database for 2018," Nucleic Acids Research, vol. 46, no. D1, 2018, pp. D1074-D1082. https://doi.org/10.1093/nar/gkx1037

[50] P. Chandak, K. Huang, and M. Zitnik, "Building a Knowledge Graph to Enable Precision Medicine," Scientific Data, vol. 10, no. 1, 2023, 67. https://doi.org/10.1038/s41597-023-01960-3

[51] Z. Ghorbanali et al., "DrugRep-KG: Toward Learning a Unified Latent Space for Drug Repurposing Using Knowledge Graphs," Journal of Chemical Information and Modeling, vol. 63, no. 8, 2023, pp. 2532-2545. https://doi.org/10.1021/acs.jcim.2c01291

[52] K. Bollacker, C. Evans, P. Paritosh, T. Sturge, and J. Taylor, "Freebase: A Collaboratively Created Graph Database for Structuring Human Knowledge," in Proceedings of the 2008 ACM SIGMOD International Conference on Management of Data (SIGMOD '08), New York, NY, USA, 2008, pp. 1247-1250. https://doi.org/10.1145/1376616.1376746

[53] G. A. Miller, "WordNet: A Lexical Database for English," Communications of the ACM, vol. 38, no. 11, 1995, pp. 39-41. https://doi.org/10.1145/219717.219748

[54] F. M. Suchanek, G. Kasneci, and G. Weikum, "Yago: A Core of Semantic Knowledge," in Proceedings of the 16th International Conference on World Wide Web (WWW '07), New York, NY, USA, 2007, pp. 697-706. https://doi.org/10.1145/1242572.1242667

[55] Bordes, N. Usunier, A. Garcia-Durán, J. Weston, and O. Yakhnenko, "Translating Embeddings for Modeling Multi-relational Data," in Proceedings of the 26th International Conference on Neural Information Processing Systems (NIPS '13), Red Hook, NY, USA, 2013, pp. 2787-2795. https://hal.science/hal-00920777

[56] Z. Sun et al., "Rotate: Knowledge Graph Embedding by Relational Rotation in Complex Space," arXiv preprint, 2019. https://doi.org/10.48550/arXiv.1902.10197

[57] B. Yang et al., "Embedding Entities and Relations for Learning and Inference in Knowledge Bases," in International Conference on Learning Representations, 2014. https://doi.org/10.48550/arXiv.1412.6575

[58] T. Trouillon et al., "Complex Embeddings for Simple Link Prediction," in Proceedings of the International Conference on Machine Learning, New York, USA, 2016, pp. 2071-2080.

[59] T. Dettmers et al., "Convolutional 2D Knowledge Graph Embeddings," in Proceedings of the AAAI Conference on Artificial Intelligence, New Orleans, USA, 2018. https://doi.org/10.1609/aaai.v32i1.11573. https://doi.org/10.1609/aaai.v32i1.11573

[60] Z. Zheng et al., "HolmE: Low-dimensional Hyperbolic KG Embedding for Better Extrapolation," in Proceedings of the ESWC Conference, 2024. https://doi.org/10.1007/978-3-031-60626-7_6

[61] A. Sarabadani et al., "DKG-LLM: A Framework for Medical Diagnosis and Personalized Treatment Recommendations via Dynamic Knowledge Graph and Large Language Model Integration," arXiv preprint, 2025. https://doi.org/10.48550/arXiv.2508.06186

[62] A. Sarabadani et al., "ExKG-LLM: Leveraging Large Language Models for Automated Expansion of Cognitive Neuroscience Knowledge Graphs," arXiv preprint, 2025. https://doi.org/10.48550/arXiv.2503.06479

[63] K. Rahsepar Fard, A. Sarabadani, and H. Dalvand, "GDPKG-LLM: Integrating Gene, Disease, and Pharmacogenomics Knowledge Graphs for Cognitive Neuroscience Using Large Language Models," csci, vol. 26, no. 3, Oct. 2025. https://doi.org/10.7494/csci.2025.26.3.6673