Journal Title Here, 2022, pp. 1–9

doi: DOI HERE
Advance Access Publication Date: Day Month Year
Paper

PAPER

An Explainable Knowledge Graph-Driven Approach to Decipher the Link Between Brain Disorders and the Gut Microbiome

Naafey Aamer,[1,*,†] Muhammad Nabeel Asim,[2,3,†] Sebastian Vollmer[1,2] and Andreas Dengel[1,2,3]

[1]Department of Computer Science, University of Kaiserslautern-Landau, Gottlieb-Daimler-Straße, Kaiserslautern, 67663, Rhineland-Palatinate, Germany, [2]DFKI, Trippstadter Str. 122, Kaiserslautern, 67663, Rhineland-Palatinate, Germany and [3] Intelligentx GmbH (intelligentx.com), Kaiserslautern, 67663, Rhineland-Palatinate, Germany
[*]Corresponding author. naafey.aamer@cs.rptu.de[†]These authors contributed equally to this work.

Abstract

Motivation: The communication between the gut microbiome and the brain, known as the microbiome-gut-brain axis (MGBA), is emerging as a critical factor in neurological and psychiatric disorders. This communication involves complex pathways including neural, hormonal, and immune interactions that enable gut microbes to modulate brain function and behavior. However, the specific mechanisms through which gut microbes influence brain function remain poorly understood, and existing computational efforts to understand these mechanisms are simplistic or have limited scope. **Results:** This work presents a comprehensive approach for elucidating the cascade of interactions that allows gut microbes to influence brain disorders. By using a large curated biomedical knowledge graph, we train GNN-GBA, an explainable graph neural network, to learn the complex biological interactions between the gut microbiome and the brain. GNN-GBA is then used to extract the mechanistic pathways through which the gut microbiome can influence brain disorders. GNN-GBA successfully identified pathways for 103 brain disorders, and we show that these pathways are consistent with existing literature. **Availability:** An interactive dashboard to explore thousands of potential mechanisms through which the gut microbiome can influence brain diseases is available at https://sds-genetic-interaction-analysis.opendfki.de/gut_brain/. Code and data are available at https://github.com/naafey-aamer/GNN-GBA. **Contact:** naafey.aamer@cs.rptu.de

Key words: gut-brain axis, knowledge graphs, brain disorders, GNNs, explainability

Introduction

Brain disorders impose a staggering global burden, they are the leading cause of disability and the second leading cause of death worldwide [18]. As shown in Figure 1, these conditions are highly complex and fall into diverse categories, including developmental, immune-mediated, neurodegenerative, and psychiatric disorders [52]. Some of these common and highly recurring conditions have no treatment options like Alzheimer's disease, Parkinson's disease, autism spectrum disorders, epilepsy, and schizophrenia [38]. The mechanisms behind brain disorders are often multifactorial, reflecting the interplay of genetic, environmental, and, as recent studies suggest [33, 47, 41, 34], microbial factors.

The relationship between brain disorders and these microbial factors is not fully understood [1], and it may be attributed to interactions between intestinal microbes and the central nervous system, which is known as the microbiome-gut-brain axis (MGBA). The MGBA is a communication network linking the gut and brain via neural, endocrine, immune and metabolic pathways [38] and a growing body of studies implicate it in neurological conditions [13, 38]. These findings suggest that the gut microbiome may be a key factor in neurodegeneration and neurodevelopment.

Central to gut-brain communication are gut metabolites, small bioactive chemicals generated during microbial metabolic processes of the digestive system [43]. These metabolites are critical messengers between the gut and the brain [43], binding to receptors on various cells in both systems to modulate key processes such as inflammation and neurotransmission [15, 7]. Beyond their role as messengers, these metabolites can influence metabolism and gene expression, resulting in functional mechanistic pathways between gut microbes and various diseases [54]. Without a clear understanding of these mechanistic pathways, our comprehension of gut microbes' role in disease remains incomplete, hindering the development of targeted therapeutic strategies. Since the gut microbiome composition is influenced by diet, medications and other confounding factors[13], existing studies often struggle to take a global view of this intricate network of cause and effect, making it challenging to develop a comprehensive understanding of

© The Author 2022. Published by Oxford University Press. All rights reserved. For permissions, please e-mail: journals.permissions@oup.com

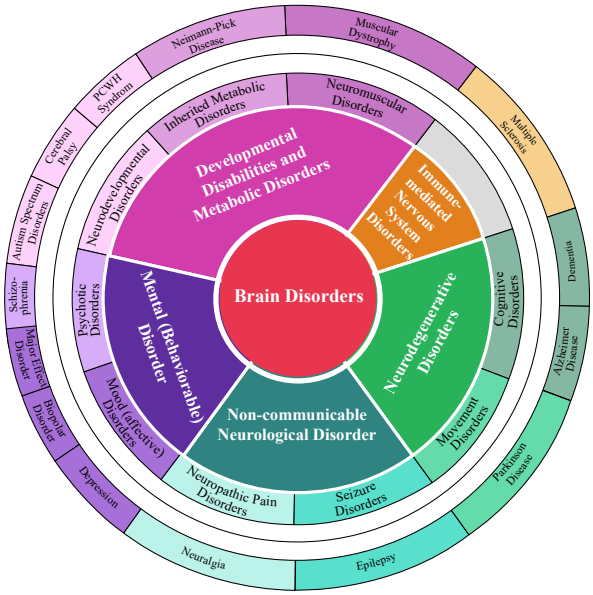

Fig. 1. Brain disorders classified into several categories and sub-categories.

these interactions. Traditional analytical methods typically examine one data modality at a time and tend to miss complex, multi-scale interactions. There is therefore a critical need for an integrative approach that can jointly analyzes diverse types of microbial, genetic, metabolic and phenotypic data.

Biomedical knowledge graphs offer a powerful solution for such integrative analysis [40]. These graphs are constructed by merging diverse data types into structured, machine-readable formats [8, 59]. They are able to represent biomedical entities (genes, proteins, microbes, diseases, metabolites, etc.) as nodes and their relationships (e.g. molecular interactions, causal links, phenotypic associations) as edges [40]. To explore the MGBA, the inclusion of metabolites is particularly crucial, as they represent the functional output of microbial activity and serve as the link between microbes and brain function. Graph machine learning has been highlighted as a key emerging tool in biomedicine, with the potential to accelerate biomedical discovery [30]. Once, the biomedical graph is constructed, graph machine learning can be directly applied to it to learn the complex interaction patterns in the MGBA.

This paper explores the mechanisms between brain disorders and the gut microbiome through knowledge graphs and advanced graph machine learning. Our contributions include: (1) a curated biomedical knowledge graph (focused on the MGBA), (2) an explainable graph neural network (GNN-GBA) that learns the underlying biological interactions within this complex graph, (3) a systematic investigation into 103 distinct brain disorders, identifying mechanistic pathways through which specific gut microbes influence these neurological conditions; and (4) we provide an interactive dashboard to explore and visualize thousands of mechanistic pathways for 103 brain disorders (link). This work not only reveals novel insights into how specific microbes might influence neurological conditions but also provides a generalizable framework for exploring other complex disease-microbiome interactions. The interactive dashboard promotes further investigation in this emerging field, offering a resource for future experimental validation and therapeutic development.

Related Work

Early studies often employed conventional machine learning on microbiome profiles – for instance, a genus-level random forest ranks microbial species by importance, but it does not explain through which biological pathway a given microbe influences a disease [51]. Such models can identify important microbial features for diseases (potential "biomarkers"), but they typically treat the microbiome as an independent high-dimensional feature vector, lacking explicit representation of mechanistic pathways. To incorporate network structure, researchers have turned to graph-based methods [37]. Since the gut-brain axis information can be efficiently represented using Knowledge Graphs (KGs), multiple works have aimed to construct them. MiKG was developed to identify potential associations between gut microbiota, neurotransmitters, and mental disorders by manually extracting relevant entities and relationships from scientific publications [32]. Additionally, established biomedical databases—such as MetaCyc [29], Reactome [12], gutMGene [10], and KEGG [28]—serve as valuable resources for KG construction, as they have collected diverse data on pathways, drugs, genes, and microbes. The Pre-/Probiotics Knowledge Graph (PPKG) integrates both scientific publications and existing biomedical databases to construct a comprehensive KG to store the relationship between the central nervous system and the enteric (gut) nervous system [31]. Beyond constructing KGs that encapsulate known relationships, there is growing interest in discovering novel MGBA pathways through predictive modeling and meta-search. Several graph based models have been proposed to predict microbe–disease links. For example, MGMLink leverages the gutMGene database to identify the shortest mechanistic pathways (i.e., gut-brain links) through breadth-first search (BFS) and template-based searches in a heterogeneous graph [49]. However, MGMLink uses node2vec [22] and TransE [4] as embedding methods which are static, precomputed, and shallow representations that do not learn global patterns within the knowledge graph, leaving a little room for comprehensive learning.

Other works employ graph neural networks (GNNs), which have shown the promise of interpretability by highlighting important microbial neighbors that drive a given disease association [19][35]. While existing GNN-based methods like [35] and [26] have potential, they lack interpretability, and the data incorporated is upto two orders of magnitude smaller than the one used in this work. This results in limited generalizability and an inability to capture the complexity of microbe-brain interactions across diverse biological pathways. To address this, this work addresses this need by training a GNN over a large curated gut–brain-axis knowledge graph of over 3.5 million relationships, encompassing microbes, their metabolites, host targets, and neurological diseases. By training on this rich network, the model learn global interaction patterns that link microbes to brain pathology. To understand these patterns, this study uses a graph interpretability method [58] that highlights relevant mechanistic pathways, effectively pointing on which intermediate nodes (specific metabolites, genes, or signaling molecules) are most critical for a predicted microbe–disease link.

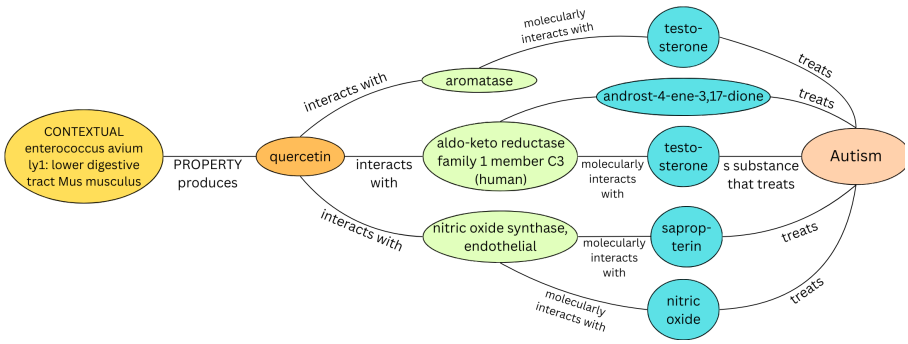

Fig. 2. A portion of knowledge graph with the detailed pathways from a microbe to a brain disorder.

Materials and Methods

This work develops a comprehensive computational framework to investigate the microbiome-gut-brain axis (MGBA) and its role in neurological disorders. The approach integrates knowledge graph construction, path mining algorithms, and graph neural networks to map the complex relationships between gut microbes, metabolites, and brain diseases. Figure 3 illustrates the overall flow of the methodology.

Knowledge Graph Construction

To create the initial MGBA knowledge graph, [48]'s methodology is utilized. First, an ontology KG was built using the PheKnowLator [6] pipeline and combined with associations from the GutMGene database. PheKnowLator is a Python 3 library that enables construction of KGs that incorporate a wide variety of data and terminology sources, including ontologies such as the Mondo Disease Ontology (MONDO) [53], the Chemical Entities of Biological Interest Ontology (CHEBI) [17], and the Human Phenotype Ontology (HPO) [21]. The gutMGene database [10] consists of microbe-metabolite and microbe-gene assertions that occur in the host of either human or mouse.

To ensure relevance and reduce noise, the constructed graph was altered to remove redundant entities and extraneous ontologies (e.g., Plant Ontology, Food Ontology, Genotype, and Transcription Factors) that did not contribute meaningfully to the scope of this work. During this process, 1,502,361 edges and 196,148 nodes were removed. The resulting knowledge graph consists of **586,318** nodes spanning 16 distinct biomedical entity types (Table 1) and **3,573,936** edges covering 103 unique predicate types. Our methodology resulted in sufficient pathways for exploring the MGBA for brain disorders, for example, Figure 2 shows the connectivity between "enterococcus avium ly1", a gut microbe, and "Autism".

It is important to note that metabolites — the critical messenger molecules between the gut and the brain — are primarily represented within the Chemical node category, which constitutes the largest node type in our graph (25.60%). This substantial representation of chemical entities, including short-chain fatty acids, neurotransmitter precursors, and bioactive peptides, enables us to trace metabolite-mediated mechanisms between gut microbes and neurological diseases. Furthermore, the relationships "molecularly interacts with," "causally influences," "causes or contributes to condition", and "is substance that treats" are particularly valuable for mapping the mechanistic pathways through which metabolites influence brain disorders.

Table 1. Node Types and Counts in the Knowledge Graph

Node Type	Count	Percentage (%)
Chemical	150,080	25.60
SNP	145,156	24.76
Protein	96,196	16.41
GO_Term	43,870	7.48
Cell_Line	41,801	7.13
Gene	26,838	4.58
Disease	23,264	3.97
Pathway	16,395	2.80
Phenotype	16,291	2.78
Anatomy	14,181	2.42
Vaccine	6,271	1.07
Taxonomy	2,373	0.40
Cell_Type	2,368	0.40
Microbe	814	0.14
Chromosome	249	0.04
Environmental_Conditions	171	0.03
Total	**586318**	**100.00**

Disease Selection And Path Mining

To get practical insights for brain disorders, it is necessary to ensure that sufficient links exist between brain disorders and the gut microbiome in the constructed knowledge graph. For that, we initially identified 2,905 diseases associated with the brain. From this large set, the focus was on diseases with robust metabolite connections, as these are critical for understanding microbe-disease relationships. Only 138 diseases had more than 30 metabolite connections, and after grouping variants of the same disease (e.g., various Alzheimer's disease subtypes) and removing conditions that returned zero paths with a timeout of 60 seconds, a final set of 103 neurological disorders were shortlisted for in-depth analysis. With these 103 diseases identified, paths between each disease and the 814 microbes in the knowledge graph were mined. For each disease-microbe pair, paths are searched, with the Networkx Python library [23], upto a maximum length of 5, capturing relevant

connections while avoiding computational overhead. Table 2 shows the top 15 brain disorders with the most disease-microbe paths. These path mining results provide the foundation for the exploration of mechanisms through which gut microbes may influence brain disorders, with metabolites playing a crucial mediating role.

Table 2. The top 15 brain disorders with the most disease-microbe paths

Brain Disorder	Interacting Metabolites	Connected Microbes	Paths
Dementia	70	288	2,029
Depression	373	274	1,808
Parkinson's	183	272	1,724
Autism	155	303	1,512
Alzheimer's	104	286	1,455
Niemann-Pick's	65	219	1,116
PCWH	65	219	1,111
Neuralgia	80	266	1,071
Multiple Sclerosis	78	215	876
Schizophrenia	133	183	847
Bipolar disorder	200	254	790
Epilepsy	172	234	789
MAD	197	152	336
Cerebral Palsy	78	121	172
Muscular Dystrophy	65	88	312

The GNN-GBA Architecture

To effectively model the complex relationships within the knowledge graph, GNN-GBA was trained for link prediction. It was implemented in PyTorch Geometric [20], a library that extends PyTorch [42] to handle graph-structured data efficiently.

Unlike traditional methods [44, 22], GNNs simultaneously capture both the local structure (direct connections) and global context (multi-hop relationships) of the biomedical entities in the graph. This is particularly crucial for understanding the gut-brain axis, where the relationships between microbes, metabolites, and brain disorders often involve complex, multi-step pathways rather than direct links.

The GNN-GBA architecture consists of a GraphSAGE [24] encoder with 3 layers. GraphSAGE employs a neighborhood sampling and aggregation strategy that learns node embeddings by sampling and aggregating features from local neighborhoods. Each layer aggregates information from neighboring nodes using mean aggregation. Batch normalization is applied after each layer (except the final layer) to stabilize training, followed by ReLU activation and dropout (p=0.3) to prevent overfitting and improve generalization. The encoder transforms node features (one-hot encoded node types) into 128-dimensional embeddings where the structural and semantic relationships in the knowledge graph are preserved.

For the decoding phase, GNN-GBA employs DistMult [57], a relation-aware scoring function that explicitly models the 103 different types of biological relationships in the graph. DistMult computes the plausibility of a triple (head entity, relation, tail entity) through bilinear scoring: $score(h, r, t) = h^T \text{diag}(r) t$, where h and t are entity embeddings and r is a learned relation-specific embedding. This relationship-aware modeling is essential for capturing the nuanced ways that microbes might influence brain disorders through metabolite-mediated mechanisms. Through the training process, the model learned to embed each entity (whether microbe, metabolite, disease, or other biomedical entity) in a high-dimensional space where the distance and direction between embeddings encode meaningful biological relationships. The key advantage of our approach over simpler alternatives is its ability to explicitly learn each relationship type separately, which enables it to differentiate between mechanistic and less relevant edges.

The model was trained using binary cross-entropy loss on both positive edges (existing relationships from the knowledge graph) and negative edges (randomly sampled non-existing relationships). The optimization was performed using Adam optimizer with an initial learning rate of 0.001 and weight decay of 5e-4. Training continued with early stopping based on validation AUC, with patience of 15 epochs to prevent overfitting.

On top of the trained GNN-GBA model, a GNNExplainer [58] module was incorporated to provide comprehensive explanation capabilities. GNNExplainer is a model-agnostic explanation framework that identifies the most important subgraph structure and node features that influence a specific prediction. The explainer learns a mask over edges by optimizing the mutual information between the GNN's prediction and the masked graph structure. It was configured with 20 training epochs and a learning rate of 0.05 to identify the pathways that most influence specific microbe-disease interactions in the knowledge graph. For each disease-microbe path, the GNNExplainer computed importance scores for individual edges within the path. The explainer generated an explanation for the overall path by creating a soft edge mask that highlights which edges in the computational graph are most critical for the model's prediction of the disease-microbe relationship. These edge-level importance scores were then aggregated using average pooling to produce a single path-level score, enabling systematic ranking of different pathways connecting the same microbe-disease pair.

Experimental Setup

To evaluate the performance of the Graph Neural Network, the knowledge graph was partitioned in a 90:5:5 ratio, resulting in 2,859,148 training samples, 357,394 validation samples, and 357,394 test samples. The validation set was used for early stopping and learning rate scheduling. Negative samples were generated through random negative sampling to create balanced datasets.

Three baseline methods combined static embeddings with machine learning classifiers. Node2Vec [22] generated 128-dimensional embeddings using walks of length 20, with 10 walks per node and context window 10, trained for 10 epochs. Edge features were constructed via Hadamard products of node embeddings and fed to Random Forests [5], XGBoost [9], and SVM [11] classifiers. Performance was evaluated using AUC-ROC, AUC-PRC, accuracy, precision, recall, and F1-score, computed both globally and stratified by relation type.

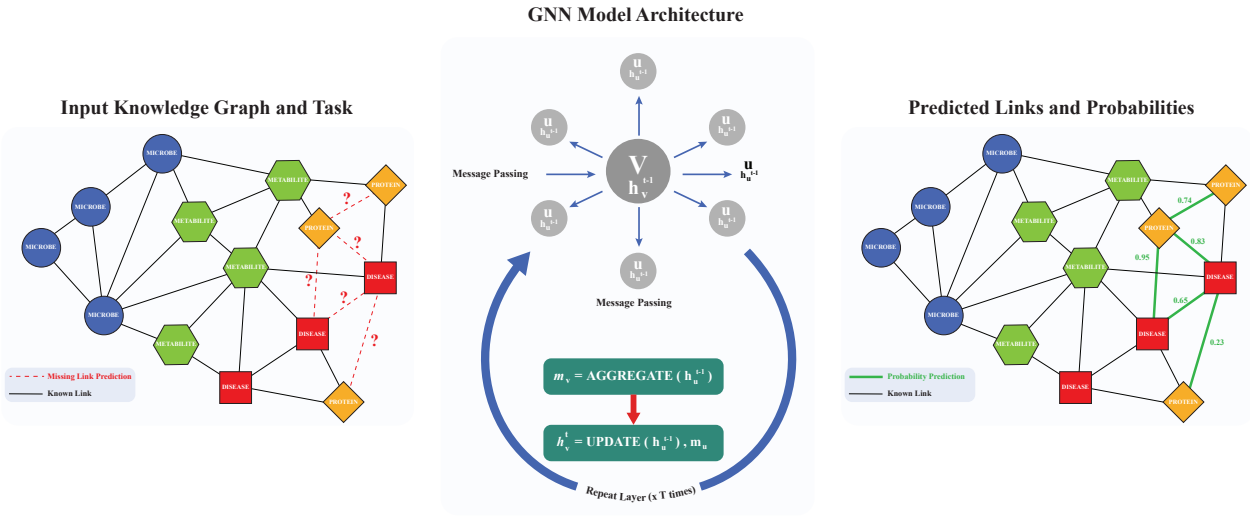

Fig. 3. The Knowledge Graph constructed using the PheKnowLater framework is into GNN-GBA which learns to predict new links.

Table 3. Relation-Wise Performance Summary Statistics

Metric	Mean	Std Dev	Min	Median	Max
AUC	0.9719	0.0476	0.7438	0.9919	1.0000
AUPRC	0.9317	0.1121	0.4688	0.9809	1.0000
F1 Score	0.7022	0.3749	0.0000	0.9333	1.0000
Precision	0.8175	0.3532	0.0000	0.9776	1.0000
Recall	0.6631	0.3892	0.0000	0.8947	1.0000
Accuracy	0.8198	0.1961	0.4500	0.9342	1.0000

Results

GNN-GBA demonstrated strong performance in predicting biological interactions within the constructed knowledge graph. As shown in Table 3, the model achieved robust relation-wise performance across 103 different relationship types, evaluated on 357,394 test samples, with a relation-wise mean of AUC of 0.9719 and mean AUPRC of 0.9317. However, the moderate mean F1 score of 0.7022, along with mean precision and recall values of 0.8175 and 0.6631 respectively, can be attributed to the class imbalance present in the knowledge graph. As illustrated in the predicate distribution, the top relationships dominate the graph structure, with "molecularly interacts with" accounting for 961,885 edges (26.9% of all edges), "subClassOf" with 614,821 edges (17.2%), and "has phenotype" with 411,610 edges (11.5%). In contrast, there are many rare relationships with fewer than 100 instances, such as "conduit for" (54 edges), "proximally connected to" (54 edges), and "in right side of" (55 edges). This imbalance explains why the model achieves excellent AUC scores (which are less sensitive to class imbalance) while showing more modest average relation-wise precision and recall metrics. The AUC metric effectively captures the model's ability to rank predictions correctly, whereas precision and recall are more affected by the underrepresentation of rare but significant relationships in the training data. Despite this challenge, the mean accuracy of 0.8198 validates the model's reliability in learning the underlying interaction patterns within the microbiome-gut-brain axis knowledge graph, particularly for the well-represented relationship types.

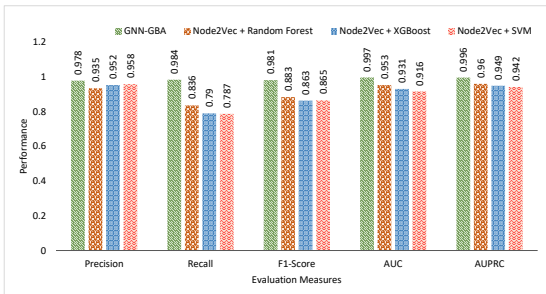

Fig. 4. Overall performance results

To validate the superiority of GNN-GBA, we compared its performance against several baseline methods including Node2Vec combined with Random Forest (N2V + RF), XGBoost (N2V + XGBoost), and Support Vector Machine (N2V + SVM). Figure 4 presents the ROC curves for all methods, demonstrating GNN-GBA's substantially outperforming the baselines. Figure 5 provides a comprehensive performance comparison across multiple metrics. These results demonstrate that GNN-GBA's can effectively capture both local structure and global context through its message-passing mechanisms.

Using the trained GNN-GBA with GNNExplainer, we identified mechanistic pathways connecting gut microbes to brain disorders. Figure 7 illustrates the top-ranked pathways for three major brain disorders: Parkinson's Disease, Alzheimer's

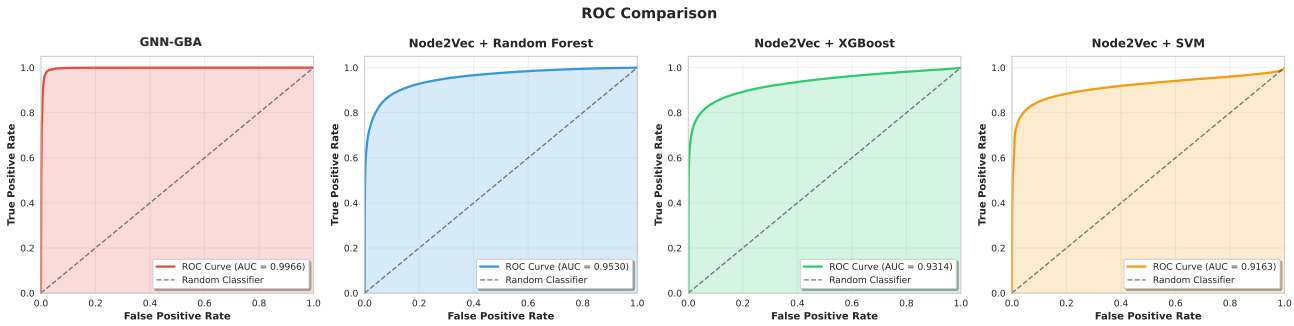

Fig. 5. AUROC Comparison with baselines methods

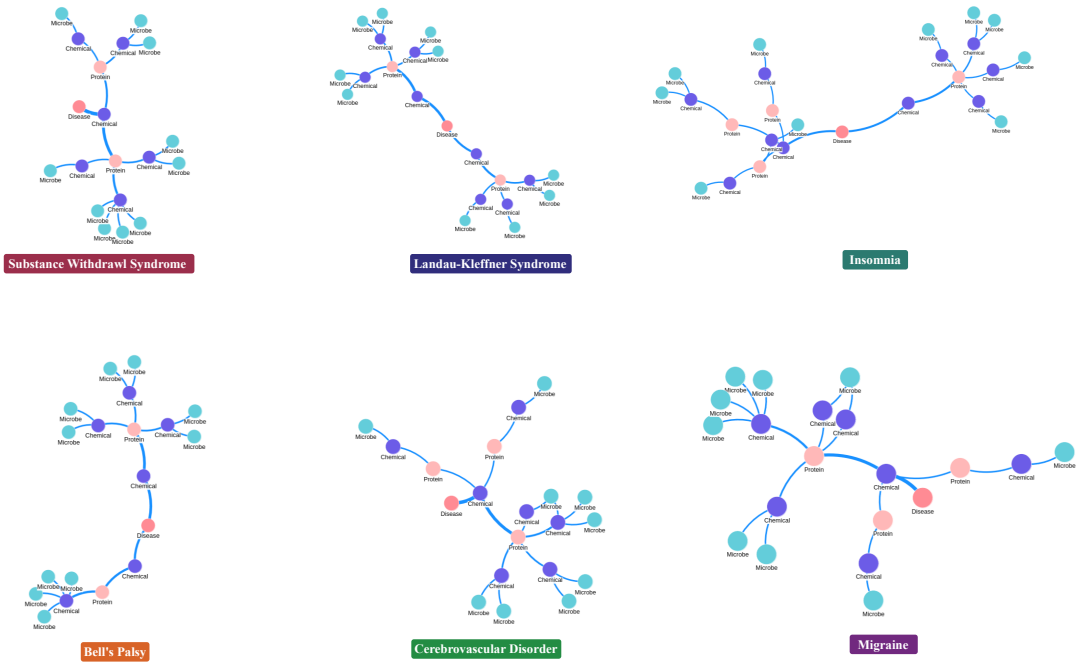

Fig. 6. Merged top paths reveal hub structures. These hubs show common pathways through which the microbiome influences neurological health.

Disease, and Major Affective Disorder. Each pathway reveals a multi-hop cascade of interactions beginning with a specific gut microbe, proceeding through metabolite and molecular interactions, and ultimately connecting to the target disease. When the top 10 pathways for each disease were merged and analyzed, interesting hub structures emerged, as visualized in Figure 6. These hub structures reveal shared mechanistic nodes that connect multiple diseases to various gut microbes, suggesting common pathways through which the microbiome influences neurological health. The presence of these hub structures also validates our approach, as they align with known biological principles where key metabolites and enzymes often participate in multiple interconnected pathways.

Brain Disorders and the Gut Microbiome

While the model's interaction prediction performance was notable, the goal is to explore the relationship of brain disorders with the gut microbiome. For each disorder, numerous microbe-disease pathways were identified. Figure 7 illustrates the top-ranked pathway for 3 common disorders as determined by GNNExplainer. Below, each of these pathways' mechanistic plausibility is verified from literature to show that practical significance of our results.

Major Affective Disorder

The top pathway connecting microbes to major affective disorder (from 336 identified paths) involves Mitsuokella multacida. Mitsuokella multacida is a anaerobic bacterium found in the lower digestive tract of humans [16] that is crucial to a healthy digestive system, Mitsuokella multacida helps unlock nutrients from plant foods (via phytate degradation) and Promotes gut and metabolic health (via propionate production). [16]. While Mitsuokella multacida itself does not directly produce creatine, it is involved in arginine regulation [14], which is a precursor for creatine synthesis [25]. Creatine is primarily synthesized in the liver from arginine, glycine, and methionine [25]. Studies suggest that creatine supplementation shows promise as a potential therapy for major affective disorders [27]. Creatine plays a crucial role in brain energy metabolism, and disruptions in this process have been linked

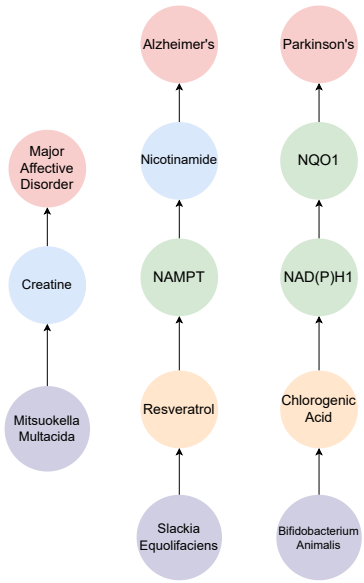

Fig. 7. Mechanistic pathways for Parkinson's Disease, Alzheimer's Disease, and Major Affective Disorder.

to mood disorders [27]. Research indicates that creatine may enhance the effectiveness of standard antidepressants such as selective serotonin reuptake inhibitors (SSRIs) [27]. This pathway suggests a biologically plausible mechanistic relationship between Mitsuokella multacida and major affective disorder through indirect influence on creatine metabolism, and warrants further investigation.

Alzheimer's Disease

The top pathway connecting microbes to Alzheimer's disease (from 1,455 identified paths) involves Slackia Equolifaciens. This bacterial species is found in the lower digestive tract of humans and is part of the gut microbiome, contributing to the overall microbial diversity of the digestive system [3]. Slackia Equolifaciens plays a key role in the metabolism of resveratrol, and has antioxidant and anti-inflammatory effects, and has also reported to mimic the beneficial effects of dietary restriction, extending the life span of mice and other species [3]. Moreover, resveratrol differentially regulates NAMPT [50], an enzyme, often activating it in healthy cells to craete NAD+(cellular energy) from nicotinamide and promote anti-aging pathways. Nicotinamide, a form of vitamin B3, has shown promise in preclinical studies for Alzheimer's disease, particularly in improving cognitive function and reducing the accumulation of amyloid plaques [45]. Studies in mouse models demonstrate that nicotinamide compounds can improve cognitive function, reduce neuroinflammation, decrease DNA damage, and enhance mitochondrial function, which is often impaired in Alzheimer's disease [2]. This pathway provides a biologically plausible mechanistic connection through which Bacillus subtilis may influence Alzheimer's disease pathology, and warrants further investigation.

Parkinson's Disease

The top pathway connecting microbes to Parkinson's disease (from 1,724 identified paths) involves Bifidobacterium Animalis. This bacterium is commonly found in the gut of humans and warm-blooded animals. B. Animalis has been shown to possess enzymatic activity capable of breaking down chlorogenic acid (CGA), a common polyphenol found in plants including coffee and green tea [46]. Studies have demonstrated that B. Animalis can utilize enzymes can break down CGA into caffeic acid, allowing for the absorption of these components in the intestines [46]. Studies show that the products of the metabolism of chlorogenic acid interact with NAD(P)H dehydrogenase [quinone] 1 (NQO1) by inhibiting its activity, regulating the enzyme's effectiveness [56, 55]. NQO1 is a crucial enzyme encoded by the NQO1 gene that plays a vital role in protecting against oxidative stress and inflammation [39]. Research suggests that dysregulation of NQO1 activity is linked to various neurological disorders, including Parkinson's disease [36]. Impaired NQO1 function may contribute to abnormal neurotransmitter release, increased oxidative stress, and cellular damage characteristic of Parkinson's disease [36]. This pathway provides a biologically plausible mechanistic pathway through which Bifidobacterium Animalis might influence Parkinson's disease pathology, and warrants further investigation.

Discussion

These pathways reveal specific mechanisms through which gut microbes may influence brain disorders. Our computationally derived mechanistic pathways align with existing literature while providing new insights into the microbiome-gut-brain axis and its interaction with brain disorders. The diversity of identified pathways across different disorders suggests condition-specific microbial influences, potentially opening avenues for targeted therapeutic approaches. The top paths for 103 distinct diseases are visualized in the GutBrainExplorer. It has an easy-to-use interface that allows researchers to select diseases from a comprehensive list and visualize the predicted pathways as interactive network graphs. These graphs display the step-by-step connections from microbes through metabolites and proteins to target diseases.

This work presents a novel approach to understanding the relationship between brain disorders and the gut microbiome using explainable graph neural networks. By integrating diverse biomedical data into a knowledge graph and applying advanced graph learning techniques, we have identified specific metabolite-mediated pathways connecting gut microbes to brain disorders. These pathways have been validated through literature and are ready for experimental validation, offering an opportunity to accelerate the development of microbe-based therapies. Moreover, these detailed pathways suggest a clear path toward developing diet recommender systems to alleviate symptoms of neurological disorders with a high global burden. Future work can entail adding dietary components to the knowledge graph and using GNN-GBA to identify dietary compounds that can modulate disease symptoms through the gut microbiome.

Competing interests

No competing interest is declared.

Author contributions statement

N.A and M.N.A conceived and conducted the experiment(s), and compiled results, and wrote the mansucript. S.V validated experiment(s) and results, and wrote the manuscript. A.D reviewed the article and performed final editing.

References

1. Cecilia Almeida, Rita Oliveira, Ricardo Soares, and Pedro Barata. Influence of gut microbiota dysbiosis on brain function: a systematic review. *Porto biomedical journal*, 5(2):e059, 2020.
2. W.J. Aronson, J.A. Glaspy, S.T. Reddy, D. Reese, D. Heber, and D. Bagga. Modulation of omega-3/omega-6 polyunsaturated ratios with dietary fish oils in men with prostate cancer. *Urology*, 58(2):283–288, 2001.
3. Lisa M Bode, Diana Bunzel, Melanie Huch, Gyu-Sung Cho, Denise Ruhland, Mirko Bunzel, Achim Bub, Charles MAP Franz, and Sabine E Kulling. In vivo and in vitro metabolism of trans-resveratrol by human gut microbiota123. *The American Journal of Clinical Nutrition*, 97(2):295–309, 2013.
4. Antoine Bordes, Nicolas Usunier, Alberto Garcia-Duran, Jason Weston, and Oksana Yakhnenko. Translating embeddings for modeling multi-relational data. In *Advances in Neural Information Processing Systems*, volume 26. Curran Associates, Inc., 2013.
5. Leo Breiman. Random forests. *Machine Learning*, 45(1):5–32, 2001.
6. Tiffany J. Callahan, Ignacio J. Tripodi, Lawrence E. Hunter, and William A. Baumgartner. A framework for automated construction of heterogeneous large-scale biomedical knowledge graphs. *bioRxiv*, 2020.
7. Marilia Carabotti, Annunziata Scirocco, Maria Antonietta Maselli, and Carola Severi. The gut-brain axis: interactions between enteric microbiota, central and enteric nervous systems. *Ann. Gastroenterol.*, 28(2):203–209, 2015. PMCID: PMC4367209.
8. P. Chandak, K. Huang, and M. Zitnik. Building a knowledge graph to enable precision medicine. *Scientific Data*, 10:67, 2023.
9. Tianqi Chen and Carlos Guestrin. Xgboost: A scalable tree boosting system. In *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, KDD '16, page 785–794. ACM, August 2016.
10. Liang Cheng et al. gutmgene: a comprehensive database for target genes of gut microbes and microbial metabolites. *Nucleic Acids Research*, 50(D1):D795–D800, 2022.
11. Corinna Cortes and Vladimir Vapnik. Support-vector networks. *Machine Learning*, 20(3):273–297, 1995.
12. David Croft, Gavin O'Kelly, Guanming Wu, Robin Haw, Marc Gillespie, and Lisa et al. Matthews. Reactome: a database of reactions, pathways and biological processes. *Nucleic Acids Research*, 39(suppl_1):D691–D697, November 2010.
13. John F. Cryan et al. The gut microbiome in neurological disorders. *The Lancet. Neurology*, 19(2):179–194, 2020.
14. Zhao-Lai Dai et al. Utilization of amino acids by bacteria from the pig small intestine. *Amino acids*, 39(5):1201–1215, 2010.
15. Boushra Dalile, Lukas Van Oudenhove, Bram Vervliet, and Kristin Verbeke. The role of short-chain fatty acids in microbiota–gut–brain communication. *Nat. Rev. Gastroenterol. Hepatol.*, 16(8):461–478, 2019.
16. Willem M De Vos et al. Phytate metabolism is mediated by microbial cross-feeding in the gut microbiota. *Nature microbiology*, 9(7):1812–1827, 2024.
17. Kirill Degtyarenko, Paula de Matos, Marcus Ennis, and Janna et al. Hastings. Chebi: a database and ontology for chemical entities of biological interest. *Nucleic Acids Research*, 36(suppl_1):D344–D350, 2008.
18. Valery L. Feigin et al. The global burden of neurological disorders: translating evidence into policy. *The Lancet. Neurology*, 19(3):255–265, 2020.
19. Junxuan Feng, Song Wu, Hui Yang, Chuan Ai, Jie Qiao, Jing Xu, and Feng Guo. Microbe-bridged disease-metabolite associations identification by heterogeneous graph fusion. *Briefings in Bioinformatics*, 23(6):bbac423, 2022.
20. Matthias Fey and Jan Eric Lenssen. Fast graph representation learning with pytorch geometric. *ICLR Workshop on Representation Learning on Graphs and Manifolds*, 2019.
21. Michael A Gargano et al. The human phenotype ontology in 2024: phenotypes around the world. *Nucleic Acids Research*, 52(D1):D1333–D1346, 2024.
22. Aditya Grover and Jure Leskovec. node2vec: Scalable feature learning for networks, 2016.
23. Aric Hagberg, Pieter Swart, and Daniel S Chult. Exploring network structure, dynamics, and function using networkx, 2008.
24. William L. Hamilton, Rex Ying, and Jure Leskovec. Inductive representation learning on large graphs. In *Proceedings of the 31st International Conference on Neural Information Processing Systems*, NIPS'17, page 1025–1035, Red Hook, NY, USA, 2017. Curran Associates Inc.
25. E. Jahangir, J. A. Vita, D. Handy, M. Holbrook, J. Palmisano, R. Beal, J. Loscalzo, and R. T. Eberhardt. The effect of l-arginine and creatine on vascular function and homocysteine metabolism. *Vascular medicine*, 14(3):239–248, 2009.
26. Cong Jiang, Junxuan Feng, Bingshen Shan, Qiyue Chen, Jian Yang, Gang Wang, Xiaogang Peng, and Xiaozheng Li. Predicting microbe-disease associations via graph neural network and contrastive learning. *Frontiers in Microbiology*, 15:1483983, 2024.
27. K. Juneja, H. P. Bhuchakra, S. Sadhukhan, I. Mehta, A. Niharika, S. Thareja, T. Nimmakayala, and S. Sahu. Creatine supplementation in depression: A review of mechanisms, efficacy, clinical outcomes, and future directions. *Cureus*, 16(10):e71638, 2024.
28. Minoru Kanehisa, Susumu Goto, Yoko Sato, Masayuki Kawashima, Miho Furumichi, and Mao Tanabe. Data, information, knowledge and principle: back to metabolism in kegg. *Nucleic Acids Research*, 42(Database issue):D199–D205, 2014.
29. Peter D. Karp, Monica Riley, Suzanne M. Paley, and Alida Pellegrini-Toole. The metacyc database. *Nucleic Acids Research*, 30(1):59–61, 2002.
30. M.M. Li, K. Huang, and M. Zitnik. Graph representation learning in biomedicine and healthcare. *Nature Biomedical Engineering*, 6:1353–1369, 2022.
31. Ting Liu, Gongjin Lan, K. Anton Feenstra, Zhisheng Huang, and Jaap Heringa. Towards a knowledge graph for pre-/probiotics and microbiota–gut–brain axis diseases.

Scientific Reports, 12(1):18977, 2022.

32. Ting Liu, Xueli Pan, Xu Wang, K. Anton Feenstra, Jaap Heringa, and Zhisheng Huang. Predicting the relationships between gut microbiota and mental disorders with knowledge graphs. *Health Information Science and Systems*, 9(1):3, 2020.
33. J. S. Loh, W. Q. Mak, L. K. S. Tan, et al. Microbiota–gut–brain axis and its therapeutic applications in neurodegenerative diseases. *Signal Transduction and Targeted Therapy*, 9:37, 2024.
34. S. K. Lotz, B. M. Blackhurst, K. L. Reagin, and K. E. Funk. Microbial infections are a risk factor for neurodegenerative diseases. *Frontiers in Cellular Neuroscience*, 15:691136, 2021.
35. Pengli Lu and Ling Li. MGDHGS: Gene-bridged metabolite-disease relationships prediction via graphsage and self-attention mechanism. *Computational Biology and Chemistry*, 108:108036, 2024.
36. Shilin Luo et al. Akt phosphorylates nqo1 and triggers its degradation, abolishing its antioxidative activities in parkinson's disease. *The Journal of Neuroscience*, 39(37):7291–7305, 2019.
37. Matthew Mayers, Ru Tu, Daniel Steinecke, T. S. Li, Núria Queralt-Rosinach, and Andrew I. Su. Design and application of a knowledge network for automatic prioritization of drug mechanisms. *Bioinformatics*, 38(10):2880–2891, 2022.
38. Mohammed M. Nakhal et al. The microbiota-gut-brain axis and neurological disorders: A comprehensive review. *Life*, 14(10):1234, 2024.
39. NCBI. NQO1 NAD(P)H quinone dehydrogenase 1 [Homo sapiens (human)]. NCBI Gene Database, 2025. Gene ID: 1728, updated on 3-May-2025.
40. David N. Nicholson and Casey S. Greene. Constructing knowledge graphs and their biomedical applications. *Computational and Structural Biotechnology Journal*, 18:1414–1428, 2020.
41. J. Park and C. H. Kim. Regulation of common neurological disorders by gut microbial metabolites. *Experimental & Molecular Medicine*, 53:1821–1833, 2021.
42. Adam Paszke, Sam Gross, Francisco Massa, Adam Lerer, and James et al. Bradbury. Pytorch: An imperative style, high-performance deep learning library. In *Advances in Neural Information Processing Systems 32*, pages 8024–8035. Curran Associates, Inc., 2019.
43. Alessia Perino and Kristina Schoonjans. Metabolic messengers: bile acids. *Nature Metabolism*, 4(4):416–423, 2022.
44. Bryan Perozzi, Rami Al-Rfou, and Steven Skiena. Deepwalk: online learning of social representations. In *Proceedings of the 20th ACM SIGKDD international conference on Knowledge discovery and data mining*, KDD '14, page 701–710. ACM, August 2014.
45. B. Puri, A.J. Richardson, D.F. Horrobin, et al. Eicosapentaenoic acid treatment in schizophrenia associated with symptom remission, normalisation of blood fatty acids, reduced neuronal membrane phospholipid turnover and structural brain changes. *International Journal of Clinical Practice*, 54(1):57–63, 2000.
46. Stefano Raimondi, Andrew Anighoro, Andrea Quartieri, Alberto Amaretti, Francisco Tomás-Barberán, Giulio Rastelli, and Maddalena Rossi. Role of bifidobacteria in the hydrolysis of chlorogenic acid. *MicrobiologyOpen*, 4:41–52, 02 2015.
47. G. Sani, M. Manchia, A. Simonetti, D. Janiri, P. Paribello, F. Pinna, and B. Carpiniello. The role of gut microbiota in the high-risk construct of severe mental disorders: A mini review. *Frontiers in Psychiatry*, 11:585769, 2021.
48. B.E. Santangelo, M. Bada, L.E. Hunter, et al. Hypothesizing mechanistic links between microbes and disease using knowledge graphs. *Scientific Reports*, 15:6905, 2025.
49. Brook Santangelo, Michael Bada, Lawrence Hunter, and Catherine Lozupone. Hypothesizing mechanistic links between microbes and disease using knowledge graphs. *bioRxiv*, December 4 2023. Preprint.
50. S. Schuster, M. Penke, T. Gorski, A. Barnikol-Oettler, R. Gebhardt, W. Kiess, and A. Garten. Resveratrol differentially regulates NAMPT and SIRT1 in hepatocarcinoma cells and primary human hepatocytes. *Cancer & Metabolism*, 2(Suppl 1):P65, 2014.
51. Qianqian Su, Qian Liu, Raymond I. Lau, et al. Faecal microbiome-based machine learning for multi-class disease diagnosis. *Nature Communications*, 13(1):6818, 2022.
52. K. T. Thakur, E. Albanese, P. Giannakopoulos, et al. Neurological disorders. In V. Patel, D. Chisholm, T. Dua, et al., editors, *Mental, Neurological, and Substance Use Disorders: Disease Control Priorities, Third Edition (Volume 4)*, chapter 5. The International Bank for Reconstruction and Development / The World Bank, Washington (DC), 2016. Available from: https://www.ncbi.nlm.nih.gov/books/NBK361950/.
53. Nicole A Vasilevsky et al. Mondo: Unifying diseases for the world, by the world. *medRxiv*, 2022.
54. Zachary D. Wallen, Ayse Demirkan, Guy Twa, and Gwendolyn et al. Cohen. Metagenomics of parkinson's disease implicates the gut microbiome in multiple disease mechanisms. *Nat. Commun.*, 13(1):6958, 2022.
55. L. Wang, X. Pan, L. Jiang, Y. Chu, and S. et al. Gao. The biological activity mechanism of chlorogenic acid and its applications in food industry: A review. *Frontiers in Nutrition*, 9:943911, 2022.
56. Shan Wang, Yue Li, Dejian Huang, Shangwei Chen, Yongmei Xia, and Song Zhu. The inhibitory mechanism of chlorogenic acid and its acylated derivatives on α-amylase and α-glucosidase. *Food Chemistry*, 372:131334, 2022.
57. Bishan Yang, Wen tau Yih, Xiaodong He, Jianfeng Gao, and Li Deng. Embedding entities and relations for learning and inference in knowledge bases, 2015.
58. Rex Ying, Dylan Bourgeois, Jiaxuan You, Marinka Zitnik, and Jure Leskovec. Gnnexplainer: Generating explanations for graph neural networks, 2019.
59. Y. Zhang, X. Sui, F. Pan, K. Yu, K. Li, S. Tian, and A. et al. Erdengasileng. A comprehensive large scale biomedical knowledge graph for ai powered data driven biomedical research. *bioRxiv: The Preprint Server for Biology*, 2025.