RESEARCH

Open Access

Mining on Alzheimer's diseases related knowledge graph to identity potential AD-related semantic triples for drug repurposing

Yi Nian[1], Xinyue Hu[1], Rui Zhang[2], Jingna Feng[1], Jingcheng Du[1], Fang Li[1], Larry Bu[4], Yuji Zhang[4], Yong Chen[3] and Cui Tao[1]*

*Correspondence:
Cui.Tao@uth.tmc.edu

[1] School of Biomedical Informatics, University of Texas Health Science Center at Houston, 7000 Fannin St, Houston, TX 77030, USA
[2] Department of Pharmaceutical Care & Health System (PCHS) and the Institute for Health Informatics (IHI), University of Minnesota, 7-115A Weaver-Densford Hall, Minneapolis, MN 55455, USA
[3] Department of Biostatistics, Epidemiology and Informatics (DBEI), the Perelman School of Medicine, University of Pennsylvania, 602 Blockley Hall, 423 Guardian Drive, Philadelphia, PA 19104, USA
[4] University of Maryland School of Medicine, 655 W Baltimore St S, Baltimore, MD 21201, USA

Abstract

Background: To date, there are no effective treatments for most neurodegenerative diseases. Knowledge graphs can provide comprehensive and semantic representation for heterogeneous data, and have been successfully leveraged in many biomedical applications including drug repurposing. Our objective is to construct a knowledge graph from literature to study the relations between Alzheimer's disease (AD) and chemicals, drugs and dietary supplements in order to identify opportunities to prevent or delay neurodegenerative progression. We collected biomedical annotations and extracted their relations using SemRep via SemMedDB. We used both a BERT-based classifier and rule-based methods during data preprocessing to exclude noise while preserving most AD-related semantic triples. The 1,672,110 filtered triples were used to train with knowledge graph completion algorithms (i.e., TransE, DistMult, and ComplEx) to predict candidates that might be helpful for AD treatment or prevention.

Results: Among three knowledge graph completion models, TransE outperformed the other two (MR = 10.53, Hits@1 = 0.28). We leveraged the time-slicing technique to further evaluate the prediction results. We found supporting evidence for most highly ranked candidates predicted by our model which indicates that our approach can inform reliable new knowledge.

Conclusion: This paper shows that our graph mining model can predict reliable new relationships between AD and other entities (i.e., dietary supplements, chemicals, and drugs). The knowledge graph constructed can facilitate data-driven knowledge discoveries and the generation of novel hypotheses.

Keywords: Alzheimer's disease, Dietary supplement, Drug repurposing, Knowledge graph, Literature mining

Background

Neurodegenerative diseases are a heterogeneous group of disorders that are characterized by the progressive degeneration of the structure and function of the central nervous system or peripheral nervous system[1]. Common neurodegenerative diseases, such as Alzheimer's disease(AD) and related dementias (ADRD), are usually incurable and irreversible and difficult to stop.

AD/ADRD are multi-factorial and complex neurodegenerative diseases characterized by progressive memory loss and severe dementia with neuropsychiatric symptoms [2]. An estimated 5.8 million Americans aged 65 and older (12.6%) are living with AD/ADRD in 2020, and this number is projected to reach 13.8 million by 2050 [3]. High prevalence of AD/ADRD creates huge medical and social burdens. The total costs for health care, long-term care and hospital services for all Americans with AD/ADRD are estimated at 305 billion in 2020 [3]. The high failure rate of the development of AD/ADRD drugs amplifies demographic and financial challenges. Given the increasing prevalence of the disease, finding innovative ways to develop effective drugs is an urgent need. Drug repurposing is a strategy for identifying new usages of approved or investigational drugs that are outside the scope of their original medical indications [4]. There are majorly three computational methods for discovering drug repurposing evidence: the network-based methods, text mining and natural language processing (NLP) based approaches, as well as machine learning-based approaches [5]. Inspired by the fact that biologic entities in the same module of biological networks share similar characteristics, network-based approach aims to find several modules(subnetworks or cliques) using algorithms according to the topology structures of networks. NLP approaches usually includes processes of identifying biological entities and mining new knowledge from scientific literature. While machine learning-based approaches can apply different machine learning models such as logistic regression, support vector machine (SVM), random forest(RF), and deep learning (DL) to identify drug repurposing signals The computational drug repurposing strategy offers various advantages over developing entirely new drugs, including the possibilities to lower failure risks and risk of unknown side effects/complications, efficient utilization of development funds and shortened development timelines [6]. Developments in high-throughput screening technologies have catapulted computational drug repurposing to the forefront of attractive drug discovery approaches because the vast amounts of available data could potentially lead to new clues for drug repurposing that individual projects could not possibly reveal.

Knowledge graphs can provide comprehensive and semantic representations for heterogeneous data, which has been successfully leveraged in many biomedical applications including drug repurposing [7]. For example, a few recent research focused on using knowledge graph-based approaches to drug repurposing for COVID-19 [8] [9] [10]. Sosa et al. applied knowledge graph embedding methods in drug repurposing for rare diseases [11]. Malas et al. leveraged the semantic properties of a knowledge graph to prioritize drug candidates for autosomal dominant polycystic kidney disease (ADPKD) [12]. However, to the best of our knowledge, knowledge graph-based approaches have rarely been applied in AD/ADRD drug repurposing.

The objective of this paper is to study potential relations between Alzheimer's diseases and dietary supplements, chemicals, and drugs using a knowledge graph-based

approach. Studies have indicated that some drugs, chemicals or food supplements could be related to preventing or delaying neurodegeneration and cognitive decline [13]. However, further research is needed to better understand the back-end mechanisms and to reveal the potential interactions with clinical and pharmacokinetic factors. In this paper, we encode biomedical concepts and their rich relations into a knowledge graph through literature mining [14]. Literature Mining is a data mining technique that identifies the entities such as genes, diseases, and chemicals from literature, discovers global trends, and facilitates hypothesis generation based on existing knowledge. Literature mining enables researchers to study a massive amount of literature quickly and reveal hidden relations between entities that were hard to be discovered by manual analysis. More specifically, we introduce a biomedical knowledge graph that specifically focuses on AD/ADRD and discovers underlying relations between chemicals, drugs, dietary supplements and AD/ADRD. More details of how to construct the knowledge graph and how to leverage graph embedding methods to predict candidates with scoring will be described in the methods section. We also present several rankings of candidates and comparisons of different graph embedding algorithms.

Results and discussion

Knowledge graph construction

There are 113,863,366 triples and 20,943,461 entities in total obtained from SemMedDB including 68 types of relations and 133 pairs of subject/object. After the rule-based filtering process described in the Preprocessing section, there are 2,811,329 triples left with a total of 128,177 subjects and objects. With further BERT-based filtering, 1,672,110 triples and 128,177 objects/subjects are left. After deduplicating triplets before training of graph embedding algorithms, there are 791,827 triples and 128,177 objects/subjects left.

Experimental settings

All 791,827 triples are split into 649,924/113,031/28,872 as training/test/validation sets respectively. The split is done in a way that we can use data from 2019 to 2020 to validate our model and triples before 2019 as the training set and triples after 2020 as the test set. Table 1 shows the performance of three widely used graph completion methods that are trained on our knowledge graph: TransE is based on translational distance and DistMult and ComplEx are based on semantic information. We can see that the TransE model performs the best among all these graph embedding algorithms with a Mean Rank (MR) of 10.53 and a Hit Ratio of 10 (Hits@10) 0.58. We then use TransE model for the prediction of potential candidates. Specifically, the final model embeds nodes into a size of 250 with a learning rate of 0.01 with an L2 distance metric.

Table 1 Graph Embedding Algorithms Performance

	MR	MRR	Hits@1	Hits@3	Hits@10
TransE	10.53	0.38	0.28	0.40	0.58
DistMult	14.58	0.23	0.13	0.21	0.40
ComplEx	12.16	0.28	0.18	0.29	0.47

Nian *et al. BMC Bioinformatics*   (2022) 23:407

Page 4 of 15

Prediction results

We found that some potential candidates might be relevant to AD prevention and treatment. Based on the training data and our scoring function, we identified the top-ranked subjects that connect with AD-related concepts with predicates *treat* or *prevent*. Tables 2,3, and 4 show the top 10 entities according to their numbers of appearances for the *drug, chemical,* and *Dietary Supplement* categories respectively. Table 5,6 and7

Table 2 Rankings For Drugs

Treat	Frequency	Prevent	Frequency
Imiquimod 50 mg/ml topical cream	4	*Amifostine*	10
Local corticosteroid injections	4	*Oral form acyclovir*	6
Oral form acyclovir	4	Betadine ointment	5
Oral form metronidazole	4	**Betadine solution**	5
Oral form prednisolone	4	*Oral form vancomycin*	5
Oral form propranolol	4	**Oxytocin injectable solution**	5
Pimecrolimus 10mg/g cream	4	Tenoxicam 20 mg	5
Topical form tacrolimus	4	**Betadine alcoholic**	4
Cam, topical lotion	3	Bromfenac ophthalmic solution	4
Dexamethasone injection	3	Corticosteroids cream	4

The triples appeared in table, with relevant evidence from PubMed with studies earlier than 1/1/2019 are marked in bold

Table 3 Rankings For Chemicals

Treat	Frequency	Prevent	Frequency
Amifostine	4	**Amoxicillin**	3
Chlorhexidine	4	**Cortisone**	3
Amiloride	2	*Fluoroquinolones*	3
Dexrazoxane	2	Streptomycin	3
Enoxaparin	2	**Allopurinol**	2
Etazolate	2	**Ampicillin**	2
Forskolin	2	Aureomycin	2
Licopyranocoumarin	2	*Ceftriaxone*	2
Local anesthesia	2	*Clarithromycin*	2
M 40403	2	Gabapentin	2

The triples appeared in table, with relevant evidence from PubMed with studies earlier than 1/1/2019 are marked in bold

Table 4 Rankings For Dietary Supplements

Affect	Frequency
Dietary fiber	8
Tea	8
Egg food product	7
Electrolytes	7
Fruit	7
Honey	7
Rice	7
Coffee	6
Ketamine	6
Acetylcysteine	5

The triples appeared in table, with relevant evidence from PubMed with studies earlier than 1/1/2019 are marked in bold

Nian *et al. BMC Bioinformatics*   (2022) 23:407

Page 5 of 15

Table 5 Rankings For Drug Triples

Treat		Prevent	
Topical form tacrolimus	*AD*	*Amifostine*	*AD*
imiquimod Topical Cream	AD	*Amifostine*	*AD, Late Onset*
Oral form prednisolone	**AD**	Betadine Ointment	AD
Pimecrolimus cream	AD	**Betadine Solution**	**AD**
Topical form tacrolimus	*AD, Early Onset*	*Oral form acyclovir*	*AD*
Oral form propranolol	**AD**	*Amifostine*	*Familial AD*
imiquimod Topical Cream	AD, Early Onset	*Amifostine*	*AD, Early Onset*
Topical form tacrolimus	*Familial AD*	**Oxytocin Injectable Solution**	**AD**
Oral form prednisolone	**AD, Early Onset**	tenoxicam 20 MG	AD
imiquimod Topical Cream	Familial AD	*Oral form acyclovir*	*AD, Late Onset*

The triples appeared in table, with relevant evidence from PubMed with studies earlier than 1/1/2019 are marked in bold

Table 6 Rankings For Chemical Triples

Treat		Prevent	
Chlorhexidine	*AD*	**Cortisone**	**AD**
ritonavir	AD	*Fluoroquinolones*	*AD*
Amifostine	**AD**	Streptomycin	AD
Tetracycline	**AD**	**Amoxicillin**	**AD**
Etazolate	**AD**	Itraconazole	AD
licopyranocoumarin	**AD**	**Pentoxifylline**	**AD**
Amiloride	**AD**	*Clarithromycin*	*AD*
Seprafilm	AD	**Ampicillin**	**AD**
propargylamine	**AD**	Penicillins	AD
Chlorhexidine	*AD*, Late Onset	Aureomycin	AD

The triples appeared in table, with relevant evidence from PubMed with studies earlier than 1/1/2019 are marked in bold

Table 7 Rankings For Dietary Supplement Triples

Affect	
Tea	*Familial AD*
Dietary Fiber	**Familial AD**
Tea	*AD 7*
Rice	**Familial AD**
Dietary Fiber	**AD**
Fruit	*Familial AD*
Egg Food Product	*Familial AD*
Tea	*AD, Late Onset*
Honey	*Familial AD*
Dietary Fiber	**AD 7**

The triples appeared in table, with relevant evidence from PubMed with studies earlier than 1/1/2019 are marked in bold

shows the top 10 ranked triples according to the candidate scores for the three categories. The triples with relevant evidence from PubMed with studies earlier than 1/1/2019 are marked in bold. The triples that only appeared in recently published studies after 1/1/2019 are marked in italic. The clinical drug and chemical categories were extracted

from the Unified Medical Language System (UMLS) and we used the Integrated Dietary Supplement Knowledge Base (iDISK) [15] as a reference for dietary supplements.

Clinical Drug

For the treatment relation, We were able to find evidence supporting seven out of ten entities (Table 2) and six out of ten triples (Table 5) through related literature and clinical trials for triples. All drugs appear in Table 5 appear in Table 2 while Table 2 has some extra drugs: Local corticosteroid, acyclovir, metronidazole, Cam, and Dexamethasone. Specifically, corticosteroids might become part of a multi-agent regimen for Alzheimer's disease and also have applications for other neurodegenerative disorders [16]. Our model indicates that Valacyclovir, an antiviral medication might also have an effect in AD/ADRD prevention. While we did not find evidence that Acyclovir is directly related to AD/ADRD, a recent study shows that Valacyclovir Antiviral therapy could be used to reduce the risk of dementia [17]. A study demonstrated that antibiotic (ABX) cocktail-mediated perturbations (high dose kanamycin, gentamicin, colistin, metronidazole, vancomycin) of the gut microbiome in two independent transgenic lines leads to a reduction in $A\beta$ deposition in male mice and underlie the observed reductions in brain amyloidosis, which is the hallmark of Alzheimer's disease.[18]. Tacrolimus [19] has been in phase two clinical trial which investigates neurobiological effect in persons with MCI and dementia starting 12/1/2021. Early study also indicated that high doses of prednisolone have the effect of reducing amyloid reduction which resulted in some delay of the cognitive decline [20][21]. Propranolol [22] has shown efficacy in reducing cognitive deficits in Alzheimer's transgenic mice. According to Joseph[16], a short pulse of high dose intrathecal methylprednisolone, dexamethasone or triamcinalone will result in detectable slowing of Alzheimer's disease.

As for the prevent relation, we found evidence that supports seven among ten triple predictions (Table 5) and all drugs in this table also appear in the Table 2. For example, a recent study in 2021 shows that Amifostine, which appears in our top 4 triple predictions, could mitigate cognitive injury induced by heavy-ion radiation [23]. Betaine could be a promising candidate for arresting Hcy-induced D-like pathological changes and memory deficits [24]. Mazurek et al. show that Oxytocin could interfere with the formation of memory in experimental animals and contribut to memory disturbance associated with Alzheimer's disease [25].

Chemical

For the treat relationship prediction, we found supporting evidence for seven out of the top ten entities (Table 3) and eight out of the top ten triple predictions (Table 6). For the treat relations, Table 3 and Table 6 have some overlaps: Amifostine, Chlorhexidine, Amiloride, Etazolate, and licopyranocoumarin. As we discussed in the Drug section, Amifostine, which appears in our top 1 triple predictions, could mitigate cognitive injury induced by heavy-ion radiation [23]. Moreover, a study has shown that oral pathogens in some circumstances can approach the brain, potentially affecting memory and causing dementia [26]. Since chlorhexidine could be used to reduce Methicillin-resistant Staphylococcus aureus (MRSA) to improve oral health, it might be a potential candidate for the treatment of Alzheimer's disease. Several studies mentioned the neuroprotective activity

of Tetracycline and its derivatives [27] [28]. Amiloride is a Na+/H+ exchangers (NHEs), which is proved to be associated with the development of mental disorders or Alzheimer's disease [29]. In addition, we found in an earlier clinical trial that Etazolate was used to moderate AD [30]. Licopyranocoumarin, as a compound from herbal medicine, was proved to have neuroprotective effect to Parkinson disease [31].

Dexrazoxane and Forskolin only appear in Table 3. A study in 2019 implies that Dexrazoxane may serve as an effective neuroprotectant to treat neurodegeneration and has potential clinical value in term of PD therapeutics[32]. Forskolin shows neuroprotective effects in APP/PS1 Tg mice and may be a promising drug in the treatment of patients with AD[33]. In addition, Tetracyline and proparglyamine only show up in Table 6. There are several studies mentioned that the neuroprotective activity of Tetracycline and its derivatives [27] [28]. Propargylamine was discussed on its beneficial effects and pro-survival/neurorescue inter-related activities relevant to Alzheimer's disease in several studies [34][35].

For prevention relation, we found six out of ten triples that are related to AD and all six corresponding chemicals also appear in Table 3. Recent studies show that antibiotic chemicals such as Fluoroquinolones, Amoxicillin, Clarithromycin, and Ampicillin can produce therapeutic effects to Alzheimer's disease [36][37]. Although we have not found that Cortisone has a direct effect on Alzheimer's disease, common anti-inflammatory drugs do have some treatment effects [38]. Earlier study has shown that allopurinol has treatment of aggressive behaviour in patients with dementia [39]. In addition, Ceftriaxone(CEF) appears in Table 3. It significantly attenuated amyloid deposition and neuroinflammatory response and a study has confirmed the potential of CEF as a promising treatment against cognitive decline from the early stages of AD progression [40].

Dietary Supplement

Since there is little evidence that food can directly treat or prevent the Alzheimer's disease, we focus on the triples with affect relationships. In the rank of the top 10 predictions of Table 4, we found dietary fiber (three times), tea (three times), rice, and honey all have the possibility to reduce the risk of AD/ADRD and they also appear in Table 7. Dietary fiber has the potential that protects impact on brain Aβ burden in older adults and the finding may assist in the development of dietary that prevent AD onset [41]. Moreover, according to [42], green tea intake might reduce the risk of dementia and cognitive impairment. Another study shows that honey can be a rich source of cholinesterase inhibitors and therefore may play a role in AD treatment [43]. Previous studies have also shown that dietary choline intake (e.g. eggs (egg yolk) and fruits) are associated with better outcomes on cognitive performance [44]. Increasing dietary intake of minerals could also reduce the risk of dementia. For example, research found a link between potassium levels and diagnosis of cognitive impairment in Mexican-Americans. [45]. In addition, one recent study indicates that highly water pressurized brown rice could ameliorate cognitive dysfunction and reduce the levels of amyloid-β, which is a major protein responsible for AD/ADRD [46]. Coffee drinking may be associated with a decreased risk of dementia/AD. This may be mediated by caffeine and/or other mechanisms like antioxidant capacity and increased insulin sensitivity.[47] Existing literature provides a reasonably strong scientific rationale to encourage testing whether ketamine (or its metabolites)

Nian *et al. BMC Bioinformatics*   (2022) 23:407

Page 8 of 15

has procognitive effects on Alzheimer's patients.[48]. Last but not least, based on the available literature, a nutraceutical formulation containing N-acetylcysteine among other compounds has shown some pro-cognitive benefits in Alzheimer's patients [49].

Conclusion

In this study, we built a framework to construct and analyze a knowledge graph that links AD/ADRD-related biomedical knowledge from PubMed to facilitate drug repurposing. More specifically, we focused on identifying potentially new relationships between AD/ADRD and chemical, drug and food supplements respectively. Our analysis indicated that the pipeline can be used to identify biomedical concepts that are semantically close to each other as well as to reveal relationships between biomedical elements and diseases of interest. Linking sparse knowledge from fast-growing literature would be beneficial for existing knowledge/information retrieval, and may promote uncovering of new knowledge. This framework is flexible and can be used for other applications such as multi-omics applications, therapeutic discovery, and clinical decision support for neurodegenerative diseases as well as other diseases. The knowledge graph we constructed can facilitate data-driven knowledge discovery and new hypothesis generation.

A breadth of possibilities exists to further improve this framework. First, our knowledge graph leveraged SemMedDB, an existing database that contains triples extracted from PubMed article. While we tried to improve the accuracy using a BERT-based approach, other NLP techniques could be implemented to further improve the accuracy of information extraction. Second, in addition to include knowledge extracted from literature, we could also incorporate triples from well-acknowledged biomedical databases to further enrich the knowledge graph. Third, we leveraged three state of the art knowledge graph embedding models in this research. In the future, we will investigate new strategies to extend embeddings to cope with sparse and unreliable data as well as multiple relationships. Last but not least, we only focused on the top 10 ranked triples for evaluation in this paper. We were able to identify supporting evidence for most of them, which indicates that our approach can inform reliable new knowledge. In addition, we only incorporate 2.8 M triples for our knowledge graph due to computational resource limits, further investigation needs to be done on additional triples, which could potentially lead to new hypotheses for AD treatment and prevention.
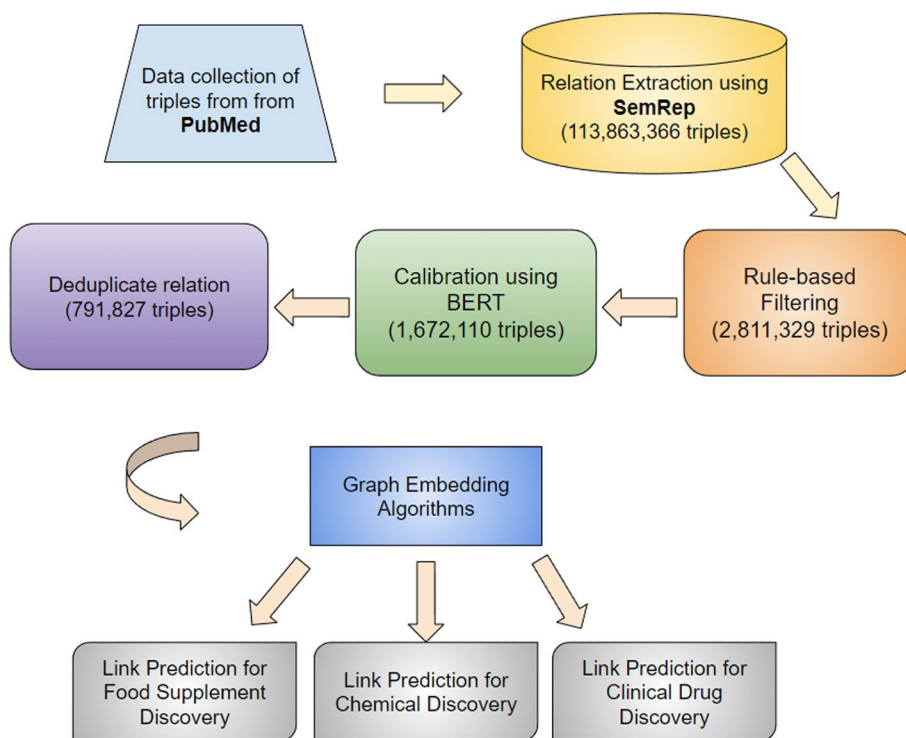
Methods

We constructed a knowledge graph using biomedical concepts and relations extracted from PubMed literature using NLP tools. The extracted triples were then further filtered based on statistics and NLP models. The rest of the subject-relation-object triples were used to build the knowledge graph. We then applied graph embedding algorithms to identify potential candidates for AD treatment and prevention. An overview of this is also described in Fig. 1.

Data Collection and Relationship Extraction

To construct the knowledge graph, we directly obtained triples from SemMedDB [50], which is a database of triples that are automatically extracted from the biomedical

Fig. 1 General Pipeline: The biological concepts in PubMed literature was extracted using NLP tools and was built into a knowledge graph using Subject-relation-object triples. Graph embedding algorithms were used to find potential candidates and complete the knowledge graph. Number of triples left are shown in each step

literature using Natural Language Processing (NLP) tools through SemRep [51]. Subject and object arguments are normalized to concepts defined in the UMLS with unique identifiers (CUIs). The triples are in the form of subject-predicate-object.

Rule-based Filtering

The original data directly obtained from the SemMedDB contained a large number of triples, but not all of them are useful for finding candidates for AD/ADRD treatment/prevention. We applied rules that are similar to [8] to exclude unrelated subject/object and predicate types. More specifically, we eliminated triples involving generic biomedical concepts such as Activities & Behaviors, Concepts & Ideas, Objects, Occupations, Organizations, and Phenomena. The rest of the triples were eliminated based on their degree of centrality (A_{in}, A_{out}) and G^2 score that indicates the strength of association between a subject and an object. Specifically, the degree centrality(A_{in}, A_{out}) was calculated with the adjacency matrix M as:

$$A_{in} = \sum_{j=1}^{n} M_{ji} \ and \ A_{out} = \sum_{j=1}^{n} M_{ij} \tag{1}$$

And the G^2 score is calculated from the statistical relation between two contingency tables: Observation table and Expectation table.[52]

$$G^2 = 2 \sum_{i,j,k} O_{ijk} * log(\frac{O_{ijk}}{E_{ijk}}) \tag{2}$$

where O_{ijk} represents the items in the observation table and

$$E_{ijk} = \frac{\sum_i O_{ijk} \sum_j O_{ijk} \sum_i O_{ijk}}{(\sum O_{ijk})^2} \tag{3}$$

represents the items in the expectation table.

At last, these three scores were normalized to [0, 1] and summed up into a final score. To keep the knowledge graph in a reasonable size that the graph embedding algorithms could handle, we only kept about 2.5 M triples. In order to ensure that AD-related triples are included in the knowledge graph, we kept all triples that are related to Alzheimer's diseases terms in the UMLS during triple elimination using the above criteria. The AD-related UMLS concepts we kept in this process were summarized in the additional file. At last, we have 2.8 M triples left in our knowledge graph.

Calibration using PubMedBERT

We leveraged about 6,000 annotations from a previous study [15] and used them as the training data for the PubMedBERT fine-tuning. These annotations were manually labeled with 1 or 0, where 1 indicates that the triples and their relationships do exist and are correct (triples labeled with 1); and 0 means that the triples do not exist or are incorrect (triples labeled with 0). PubMedBERT took the text input of subject, object, predicate type as well as the sentence that these were extracted from. The model obtained an F-1 score of 0.82, Recall of 0.91 and Precision of 0.75 on the validation set; and F-1 score of 0.83, Recall of 0.89 and Precision of 0.78 on these annotations.

Graph Embedding Algorithms

Knowledge graph embedding is a promising approach to graph completion tasks [53]. It embeds entities and relations into vector space to evaluate the probability that a given triplet (h,r,t) is true through a scoring function. We leveraged three popular knowledge graph embedding methods, TransE, DistMult and ComplEx for our knowledge graph completion task. To train this knowledge graph, these three models do negative sampling by corrupting triplets (h,r,t) to either form (h',r,r) or (h,r,t'), where h' and t' are the negative samples. Therefore, if y=±1 is the label for positive and negative triplets and f is the scoring function, then the logistic loss is computed as according to [54]:

$$\sum_{(h,r,t) \in D^+ \cup D^-}^{n} log(1 + e^{-y*f(h,r,t)}) \tag{4}$$

Tables

TransE

TransE [55] is one of the earliest translational distance models. The model projects head, tail and relations into the same space where the relation is interpreted as a translation

Nian *et al. BMC Bioinformatics*   (2022) 23:407

Page 11 of 15

vector r so that the head and tail can be connected by relations with low error. And the score function is the negative of the distance of this error as shown in Table 8. TransE does have disadvantages in dealing with 1-to-N, N-to-1, and N-to-N relations. For example, if Alzheimer's disease could be affected by different food supplements, then TransE model might learn similar results for all these food supplements.

DistMult

Semantic matching models like DistMult[56] use similarity-based scoring functions that associate each entity with a vector to capture its latent semantics. In this model, each relation is represented as a diagonal matrix which models pairwise interactions between latent factors by a bilinear function as shown in Table 8.

ComplEx

Since the scoring function of DistMult is symmetric in terms of h and t, the function cannot handle asymmetric relationships. Complex Embeddings (ComplEx) [57] introduces complex-valued embeddings to solve this problems. Specifically, the scoring function can be expanded as:

$$Re(h^T(diag(r))\bar{t} = Re(\sum_{i=0}^{d-1}[r]_i[r]_i[\bar{t}]_i) \tag{5}$$

Candidates scoring for repurposing

We focused on three kinds of predictions for the candidate selection in this research: dietary supplements candidates, chemical candidates, and clinical drug candidates. The clinical drug and chemical categories were extracted from the UMLS and we used the iDISK [15] as a reference for dietary supplements. For each type of candidates, the model iterates over all possible triples, i.e. (h_i, r_i, t_k), and $h_i \in$ all nodes for particular type of candidates, $r_j \in$ all relations, and $t_k \in$ all nodes related to Alzheimer's disease. In knowledge graph embedding-based approaches, the scoring function ϕ(h, r, t) is defined in terms of the embeddings of entities and relations; i.e., h, r, and t are embedded into vector space, and ϕ is defined in terms of operations or scoring functions over these objects. They all project the node and entities to lower-dimensional embeddings but with different scoring functions. TransE simply uses the distance between the embeddings of the head, sum with the relation embedding and tail as the scoring function, while DistMult and ComplEx use bilinear map to define scoring functions. For drugs and chemicals, we used two

Table 8 Scoring Function of Graph Embedding Algorithms

Model	Scoring type	Scoring function
TransE	Distance function	$-\|\mathbf{h} + \mathbf{r} - \mathbf{t}\|$
ComplEx	Bilinear function	$Re(h^T(diag(r))\bar{t})$
DistMult	Bilinear function	$h^T Re(diag(r))t$

The triples appeared in table, with relevant evidence from PubMed with studies earlier than 1/1/2019 are marked in bold

types of relations (i.e., treat and prevent) for prediction in this paper since the focus of the paper is drug repurposing. For dietary supplements, on the other hand, we focus on the "affect" relationship since it might be relatively challenging to detect top-ranked direct relationships between dietary supplements and AD treatment/prevention (Additional file 1: Table S9).

Evaluation for drug repurposing

We leveraged the time-slicing technique that is commonly used in literature mining [58] to evaluate our triple prediction approach. We trained all three models using data before 1/1/2019 to see whether we can predict triples that were first published after this date.

Abbreviations

AD	Alzheimer's disease
MR	Mean rank
Hits@1/3/10	Hit ratio at one/three/ten
ADRD	Alzheimer's disease and related dementias
NLP	Natural language processing
SVM	Support vector machine
RF	Random forest
DL	Deep learning
ADPKD	Autosomal dominant polycystic kidney disease
MRR	Mean reciprocal rank
iDISK	Integrated dietary supplement knowledge base
ABX	Antibiotic
MRSA	Methicillin-resistant *Staphylococcus aureus*
NHEs	Na+/H+ exchangers
CEF	Ceftriaxone
CUIs	UMLS concept unique identifiers
ComplEx	Complex embeddings
UMLS	Unified medical language system

Supplementary information

The online version contains supplementary material available at (https://doi.org/10.1186/s12859-022-04934-1).

Additional file 1. AD-related UMLS concepts used in the study.

Author Contributions
CT conceived the research project. YN, JD and CT designed the pipeline and method. YN implemented the deep learning model of the study and prepared the manuscript. JF and FL conducted the result interpretation. XH and LB prepared the data and proceed the pipeline. RZ, YC, and YZ provided expertise and suggestions on data filtering and model design especially for the dietary supplement data. All authors proofread the paper and provided valuable suggestions. All the authors have read and approved the final manuscript.

Availability of data and materials
SemMedDB: https://lhncbc.nlm.nih.gov/ii/tools/SemRep_SemMedDB_SKR.html UMLS: https://www.nlm.nih.gov/research/umls/index.html iDISK:https://academic.oup.com/jamia/article/27/4/539/5740032?login=true

Nian *et al. BMC Bioinformatics*   (2022) 23:407

Page 13 of 15

Declarations

References

1. Neurodegenerative diseases latest research and news. https://www.nature.com/subjects/neurodegenerative-diseases. Accessed 01 Oct 2022.
2. Moya-Alvarado G, Gershoni-Emek N, Perlson E, Bronfman F. Neurodegeneration and Alzheimer's disease (ad). What can proteomics tell us about the Alzheimer's brain? Mol Cell Proteom. 2016;15(2):409–25.
3. Duan R, Boland M, Liu Z, Liu Y, Chang H, Xu H, Chu H, Schmid C, Forrest C, Holmes J, Schuemie M, Berlin J, Moore J, Chen Y. Learning from electronic health records across multiple sites: a communication-efficient and privacy-preserving distributed algorithm. J Am Med Inf Assoc. 2020;27(3):376–85.
4. Ashburn T, Thor K. Drug repositioning: identifying and developing new uses for existing drugs. Nat Rev Drug Discov. 2004;3:673–83.
5. Kyungsoo P. A review of computational drug repurposing. Trans Clin Pharmacol. 2019;27(2):59–63. https://doi.org/10.12793/tcp.2019.27.2.59.
6. Pushpakom S, Iorio F, Eyers P, Escott K, Hopper S, Wells A, Doig A, Guilliams T, Latimer J, McNamee C, Norris A, Sanseau P, Cavalla D, Pirmohamed M. Drug repurposing: progress, challenges and recommendations. Nat Rev Drug Discov. 2019;18(1):41–58.
7. Bonner S, Barrett I, Ye C, Swiers R, Engkvist O, Bender A, Hoyt C, Hamilton W. A review of biomedical datasets relating to drug discovery: a knowledge graph perspective. arXiv preprint arXiv:2102.10062 (2021)
8. Zhang R, Hristovski D, Schutte D, Kastrin A, Fiszman M, Kilicoglu H. Drug repurposing for COVID-19 via knowledge graph completion. J Biomed Inf. 2021;115: 103696.
9. Yan V, Li X, Ye X, Ou M, Luo R, Zhang Q, Tang B, BJ C, I H, Siu C, ICK W, RCK C, EW C. Drug repurposing for the treatment of COVID-19: a knowledge graph approach. Adv Ther (Weinh). 2021;4(10):2100179.
10. Al-Saleem J, Granet R, Ramakrishnan S, Ciancetta N, Saveson C, Gessner C, Zhou Q. Knowledge graph-based approaches to drug repurposing for COVID-19. J Chem Inf Model. 2021;61(8):4058–67. https://doi.org/10.1021/acs.jcim.1c00642 (**PMID: 34297570**).
11. Sosa, D, Derry A, Guo M, Wei E, Brinton C, Altman R. A literature-based knowledge graph embedding method for identifying drug repurposing opportunities in rare diseases. Pac Symp Biocomput 2020; 25
12. Malas T, Vlietstra W, Kudrin R, Starikov S, Charrout M, Roos M, Peters D, Kors J, Vos R, PAC H, Mulligen E, Hettne K. Drug prioritization using the semantic properties of a knowledge graph. Sci Rep. 2019;9(1):6281.
13. Joseph J, Cole G, Head E, Ingram D. Nutrition, brain aging, and neurodegeneration. J Neurosci. 2009;29:12795–801.
14. PubMed. https://pubmed.ncbi.nlm.nih.gov/. Accessed 2022.
15. Rizvi R, Vasilakes J, Adam T, Melton G, Bishop J, Bian J, Tao C, Zhang R. iDISK: the integrated DIetary supplements knowledge base. J Am Med Inf Assoc. 2020;27(4):539–48.
16. Alisky JM. Intrathecal corticosteroids might slow Alzheimer's disease progression. Neuropsychiatr Dis treat. 2008;45:831.
17. Devanand D, Andrews H, Kreisl W, Razlighi Q, Gershon A, Stern Y, Mintz A, Wisniewski T, Acosta E, Pollina J, Katsikoumbas M, Bell K, Pelton G, Deliyannides D, Prasad K, Huey E. Antiviral therapy: valacyclovir treatment of Alzheimer's disease (valad) trial: protocol for a randomised, double-blind, placebo-controlled, treatment trial. BMJ Open. 2020;10(2): e0321112.
18. Dodiya H, Frith M, Sidebottom A, Cao Y, Koval J, Chang E, Sisodia S. Synergistic depletion of gut microbial consortia, but not individual antibiotics, reduces amyloidosis in appps1-21 Alzheimer's transgenic mice. Sci Rep. 2020;10:1–10.
19. A pilot open labeled study of tacrolimus in Alzheimer's disease. https://clinicaltrials.gov/ct2/show/results/NCT04263519. Accessed 2022.
20. Alisky J. Intrathecal corticosteroids might slow Alzheimer's disease progression. Neuropsychiatr Dis Treat. 2008;4(5):831.
21. Ricciarelli R, Fedele E. The amyloid cascade hypothesis in Alzheimer's disease: it's time to change our mind. Curr Neuropharmacol. 2017;15(6):926–35.
22. Dobarro M, Gerenu G, Ramírez M. Propranolol reduces cognitive deficits, amyloid and tau pathology in Alzheimer's transgenic mice. Int J Neuropsychopharmacol. 2013;16(10):2245–57.
23. Boutros S, Zimmerman B, Nagy S, Lee J, Perez R, Raber J. Amifostine (wr-2721) mitigates cognitive injury induced by heavy ion radiation in male mice and alters behavior and brain connectivity. Front Physiol. 2021. https://doi.org/10.3389/fphys.2021.770502.
24. Chai G-S, Jiang X, Ni Z-F, Ma Z-W, Xie A-J, Cheng X-S, Wang Q, Wang J-Z, Liu G-P. Betaine attenuates Alzheimer-like pathological changes and memory deficits induced by homocysteine. J Neurochem. 2013;3:388–96.
25. Mazurek M, Beal M, Bird E, Martin J. Oxytocin in Alzheimer's disease: postmortem brain levels. Neurology. 1987;37(6):1001–1001.

26. Sureda A, Daglia M, Castilla S, Sanadgol N, Nabavi S, Khan H, Belwal T, Jeandet P, Marchese A, Pistollato F, Forbes-Hernandez T, Battino M, Berindan-Neagoe I, G D, Nabavi S. Oral microbiota and Alzheimer's disease: Do all roads lead to rome?. Pharmacol Res. 2020;151:104582.

27. Li C, Yuan K, Schluesener H. Impact of minocycline on neurodegenerative diseases in rodents: a meta-analysis. Rev Neurosci. 2013;24(5):553–62.

28. Bortolanza M, Nascimento G, Socias S, Ploper D, Chehín R, Raisman-Vozari R, Del-Bel E. Tetracycline repurposing in neurodegeneration: focus on Parkinson's disease. J Neural Transm (Vienna). 2018;125(10):1403–15.

29. Verma V, Bali A, Singh N, Jaggi A. Implications of sodium hydrogen exchangers in various brain diseases. J Basic Clin Physiol Pharmacol. 2015;26(5):417–26.

30. A study to determine the clinical safety/tolerability and exploratory efficacy of EHT 0202 as adjunctive therapy to acetylcholinesterase inhibitor in mild to moderate Alzheimer's disease (EHT0202/002). https://clinicaltrials.gov/ct2/show/NCT00880412. Accessed 2022.

31. Fujimaki T, Saiki S, Tashiro E, Yamada D, Kitagawa M, Hattori N, Imoto M. Identification of licopyranocoumarin and glycyrurol from herbal medicines as neuroprotective compounds for parkinson's disease. PLoS One. 2014;9(6): e100395.

32. Mei M, Zhou Y, Liu M, Zhao F, Wang C, Ding J, Lu M, Hu G. Antioxidant and anti-inflammatory effects of dexrazoxane on dopaminergic neuron degeneration in rodent models of parkinson's disease. Neuropharmacology. 2019;160: 107758. https://doi.org/10.1016/j.neuropharm.2019.107758.

33. Owona B, Zug C, Schluesener H, Zhang Z. Protective effects of Forskolin on behavioral deficits and neuropathological changes in a mouse model of cerebral amyloidosis. J Neuropathol Exp Neurol. 2016. https://doi.org/10.1093/jnen/nlw043.

34. Carreiras M, Ismaili L, Marco-Contelles J. Propargylamine-derived multi-target directed ligands for Alzheimer's disease therapy. Bioorg Med Chem Lett. 2020;30(3): 126880.

35. Amit T, Bar-Am O, Mechlovich D, Kupershmidt L, Youdim MOW. The novel multitarget iron chelating and propargylamine drug m30 affects app regulation and processing activities in Alzheimer's disease models. Neuropharmacology. 2017;123:359–67.

36. Ou H, Chien W, Chung C, Chang H, Kao Y, Wu P, Tzeng N. Association between antibiotic treatment of chlamydia pneumoniae and reduced risk of Alzheimer dementia: a nationwide cohort study in taiwan. Front Aging Neurosci. 2021. https://doi.org/10.3389/fnagi.2021.701899.

37. Angelucci F, Cechova K, Amlerova J, Hort J. Antibiotics, gut microbiota, and Alzheimer's disease. J Neuroinflammation. 2019; 16108

38. Jaturapatporn D, Isaac M, McCleery J, Tabet N. Aspirin, steroidal and non-steroidal anti-inflammatory drugs for the treatment of Alzheimer's disease. Cochrane Database Syst Rev. 2012;2:CD006378.

39. Lara D, Cruz MR, Xavier F, Souza D, Moriguchi E. Allopurinol for the treatment of aggressive behaviour in patients with dementia. Int Clin Psychopharmacol. 2003;18:53–5.

40. Tikhonova MA, Amstislavskaya TG, Ho Y, Akopyan AA, Tenditnik MV, Ovsyukova MV, Bashirzade AA, Dubrovina NI, Aftanas LI. Neuroprotective effects of ceftriaxone involve the reduction of aβ burden and neuroinflammatory response in a mouse model of Alzheimer's disease. Front Neurosci. 2021. https://doi.org/10.3389/fnins.2021.736786.

41. Fernando W, Stephanie R, Gardener S, Villemagne V, Burnham S, Macaulay SL, Brown B, Gupta VB, Sohrabi H, Weinborn M, Taddei K, Laws S, Goozee K, Ames D, Fowler C, Maruff P, Masters C, Salvado O, Rowe C, Martins R. Associations of dietary protein and fiber intake with brain and blood amyloid-β. J Alzheimers Dis. 2018;61(4):1589–98.

42. Kakutani S, Watanabe H, Murayama N. Green tea intake and risks for dementia, Alzheimer's disease, mild cognitive impairment, and cognitive impairment: a systematic review. Nutrients. 2019. https://doi.org/10.3390/nu11051165.

43. Baranowska-Wójcik E, Szwajgier D, Winiarska-Mieczan A. Honey as the potential natural source of cholinesterase inhibitors in Alzheimer's disease. Plant Foods Human Nutr. 2020;75(1):30–2.

44. Ylilauri M, Voutilainen S, Eija L, Virtanen HEK, Tuomainen T, Salonen J, Virtanen J. Associations of dietary choline intake with risk of incident dementia and with cognitive performance: the Kuopio Ischaemic heart disease risk factor study. Am J Clin Nutr. 2019;110:1416–23.

45. Vintimilla RM, Large SE, Gamboa A, Rohlfing GD, O'Jile JR, Hall JR, O'Bryant SE, Johnson LA. The link between potassium and mild cognitive impairment in Mexican-Americans. Dement Geriatr Cogn Dis Extra. 2018. https://doi.org/10.1159/000488483.

46. Okuda M, Fujita Y, Katsube T, Tabata H, Yoshino K, Hashimoto M, Sugimoto H. Highly water pressurized brown rice improves cognitive dysfunction in senescence-accelerated mouse prone 8 and reduces amyloid beta in the brain. BMC Complement Altern Med. 2018;68(1):110.

47. Eskelinen MH, Kivipelto M. Caffeine as a protective factor in dementia and Alzheimer's disease. J Alzheimers Dis. 2010. https://doi.org/10.3233/JAD-2010-1404.

48. Smalheiser NR. Ketamine: a neglected therapy for Alzheimer disease. Front Aging Neurosci. 2019. https://doi.org/10.3389/fnagi.2019.00186.

49. Hara Y, McKeehan N, Dacks PA, Fillit HM. Evaluation of the neuroprotective potential of n-acetylcysteine for prevention and treatment of cognitive aging and dementia. J Prev Alzheimers Dis. (2017)

50. Kilicoglu H, Shin D, Fiszman M, Rosemblat G, Rindflesch T. Semmeddb: a pubmed-scale repository of biomedical semantic predications. Bioinformatics. 2012;28(23):3158–60.

51. Kilicoglu H, Rosemblat G, Fiszman M, Shin D. Broad-coverage biomedical relation extraction with SemRep. BMC Bioinform. 2020;21:1–28.

52. McInnes BT. Extending the log-likelihood measure to improve collocation identification. Master's thesis, Univerity of Minnesota, Minneapolis (2004)

53. Lin Y, Liu Z, Sun M, Liu Y, Zhu X. Learning entity and relation embeddings for knowledge graph completion. In: Proceedings of the Twenty-Ninth AAAI conference on artificial intelligence, AAAI'15, AAAI Press, Austin, Texas, 2015; pp. 2181–2187.

54. Zheng D, Song X, Ma C, Tan Z, Ye Z, Dong J, Xiong H, Zhang Z, Karypis G. Dgl-ke: Training knowledge graph embeddings at scale. In: Proceedings of the 43rd international ACM SIGIR conference on research and development in information retrieval. SIGIR '20, Association for computing machinery, New York, NY, USA, 2020; pp. 739–748.

55. Bordes A, Usunier N, Garcia-Duran A, Weston J, Yakhnenko O. Translating embeddings for modeling multi-relational data 2013; 26

56. Yang B, Yih W, X H, Gao J, Deng L. Embedding entities and relations for learning and inference in knowledge bases (2015). arXiv:1412.6575

57. Trouillon T, Welbl J, Riedel S, Gaussier E, Bouchard G. Complex embeddings for simple link prediction (2016). arXiv:1606.06357

58. Henry S, McInnes BT. Literature based discovery: models, methods, and trends. J Biomed Inf. 2017;74:20–32. https://doi.org/10.1016/j.jbi.2017.08.011.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.