Journal of Nonlinear and Convex Analysis
Volume 26, Number 7, 2025, 2401–2413

Yokohama Publishers
ISSN 1880-5221 ONLINE JOURNAL

# CONSTRUCTION OF A BILINGUAL KNOWLEDGE GRAPH FOR CORONAVIRUS BASED ON LLMS

PEI LOU*, AN FANG, KUANDA YAO, WANQING ZHAO, CHENLIU YANG, AND JIAHUI HU†

Abstract. Background: With the outbreak and global pandemic of the COVID-19, a large number of Chinese and English literature and data have been produced in related fields. These resources are of great value to researchers and medical workers, but there are difficulties in linking and integrating knowledge due to language differences and information fragmentation. It is critical to build a high-quality bilingual knowledge graph of coronavirus to integrate multi-source knowledge. Objective: A bilingual knowledge fusion approach based on large language model (LLMs) was proposed to construct a bilingual knowledge graph for coronaviruses, and the graph was used to discover potential targets and drug repurposing. Methods: The Llama2-Chinese-13b-Chat was selected as the basic LLM, and Lora method was used to fine-tune the model for bilingual entity translation. Then, a similarity calculation method of recalling and sorting entities was proposed for data fusion to construct the bilingual coronavirus knowledge graph. Results: The fine-tuned LLM achieved a translation accuracy of 73.72% on the test set, which was higher than the 45.78% accuracy using open source translation software. Finally, 4330 entities were fused. Through semantic retrieval and utilization, five potential drugs and three targets related to the process of coronavirus infection were identified based on the constructed bilingual knowledge graph. Conclusions: In this study, a cross-language entity alignment method based on LLMs was studied. The constructed bilingual knowledge graph of coronavirus can provide richer and high-quality content and facilitating advancements in biomedicine.

## 1. Introduction

With the development of globalization, it has become a trend to acquire, exchange and apply knowledge in multilingual environment. Cross-language knowledge graph can integrate knowledge resources of different languages, provide unified knowledge representation in multi-language environment, and meet the diverse knowledge needs of users [28]. Moreover, the cross-language knowledge graph can make full use of the complementarity of multi-language knowledge, make up for the deficiency of a single language knowledge base, and provide more comprehensive

2020 *Mathematics Subject Classification.* 68-11, 68T01.

*Key words and phrases.* Data fusion, LLMs, bilingual knowledge graph, coronavirus, knowledge discovery.

*The first author was partially supported by the CAMS Innovation Fund for Medical Sciences (CIFMS) (Grant No. 2021-I2M-1-056) and the National Social Science Foundation of China (Grant No. 21CTQ016).

†Corresponding author.

knowledge. Furthermore, the cross-language knowledge graph can integrate knowledge resources of multiple languages, cross-validate and filter information, reduce false information, and improve the accuracy and reliability of knowledge [8].

The task of cross-language entity alignment is an important part of constructing multilingual knowledge graph. Cross-language entity alignment methods are mainly divided into two categories: traditional character matching methods and methods based on deep learning [18]. Traditional methods translate bilingual entities and then align them. The accuracy of this method is limited by the quality of machine translation and needs to deal with issues such as translation ambiguity. Methods based on deep learning require training on a large amount of annotated data [2]. There are a large number of uncommon entities in the medical field, and data annotation consumes a lot of manpower. Large language models can capture more rich and complex language representations through large parameter scale and deep neural network structure [7]. Compared with traditional methods, LLMs can effectively capture the subtle differences in language, so as to understand natural language more accurately. In addition, large language models are pre-trained on a large amount of data, which have strong generalization ability and can adapt to the needs of different fields and tasks. This flexibility enables models to be easily migrated to new application scenarios without extensive domain adaptation effort [20].

The global outbreak of the COVID-19 has significantly impacted human daily life and work. coronavirus such as severe acute respiratory syndrome coronavirus (SARS-CoV), Middle East respiratory syndrome coronavirus (MERS-CoV), and SARS-CoV-2, among others, have caused numerous large-scale fatalities and sparked global panic in this century due to their high pathogenicity [11, 23]. Numerous Chinese and English literatures and data related to coronavirus have been accumulated in relevant research fields, offering significant academic and practical value for researchers and medical professionals. However, with the ongoing evolution of coronaviruses, there is an urgent need to overcome language barriers and effectively integrate scattered knowledge to facilitate in-depth study of these viruses. This paper aims to extract pertinent knowledge about coronaviruses from 36,340 Chinese and English articles and propose a cross-language entity fusion method based on LLMs. The constructed knowledge graph is expected to provide more comprehensive and systematic drug and target discovery information for researchers, thereby promoting the further development of coronavirus research.

## 2. Related work

As globalization accelerates, the study of cross-lingual knowledge graph entity alignment has garnered significant attention. Entity alignment constitutes a crucial task within the realm of natural language processing. The objective of cross-lingual entity alignment is to identify identical or analogous entities across different languages' knowledge graphs and subsequently align them. Nonetheless, due to disparities in semantics, structure, and grammar among languages, executing cross-lingual entity alignment poses considerable challenges.

The conventional approach to cross-language entity alignment involves directly translating the entity name from the source language to the target language, and subsequently aligning it with the corresponding entity in the target language. Fu et al. [9,10] introduced a comprehensive framework for this purpose. In their method, machine translation tools are employed to convert entities from one language to another, followed by a monolingual alignment technique to identify aligned entity pairs. The efficacy of this strategy is significantly influenced by the quality of machine translation. Spohr et al. [22] adopted machine learning techniques to align entities using machine-translated tags and explored multi-language entity alignment. The results showed that co-translating multi-language entities into an intermediary language can reduce the impact of machine translation, thereby enhancing experimental outcomes.

Cross-language entity alignment based on deep learning mainly focuses on how to achieve entity alignment between different languages through machine learning methods. Some researchers propose methods based on semantic and structural information to improve the accuracy of entity alignment, other methods use contextual information for entity alignment, and some methods integrate attribute embedding and relational attention, and use machine translation models, etc. way to align entities.

Kang et al. proposed a cross-language entity alignment model [12] that combines knowledge graphs and entity description information. Initially, this model employs TransC and parameter sharing models to map all entities and relationships within the knowledge graph into a shared low-dimensional semantic space of entities based on alignment. Then, model iteration and soft alignment strategies are performed to perform entity alignment. Experimental results show that the proposed model can effectively fuse ontology information and achieve better results. Zhao et al. [30] proposed a cross-language entity alignment method based on graph convolutional neural network and graph attention network. By employing multi-level learning of entity structure, attributes, and attention, it assigns appropriate weights to neighboring nodes of varying nodes, thereby capturing extensive spatial information. Zhang et al. [29] proposed a method to minimize adjacent entity filtering rules by integrating entity names and attributes (NENA). This method utilizes NENA filtering rules to filter out redundant equivalent entities and construct a dual-relation graph as auxiliary evidence for scenarios when the attribute information may be insufficient. The product network embeds the knowledge graph and entity names into a unified vector space, then applies a down sampling technique to extract the sub-graphs of the knowledge graph. This sub-graph is embedded into GCN as a new input. The cross-graph-matching module is subsequently employed to achieve alignment. Wang et al. [26] introduced the FuAlign model, a novel cross-lingual entity alignment framework based on multi-view knowledge representation learning of a pre-fused knowledge graph. FuAlign first fuses two matching knowledge graphs based on the given seed set. Then, it exploits multi-view representation learning to map the fused knowledge graph into a unified space. This multi-view representation learning approach is adept at capturing diverse information types, including semantics, entity context, and long-term entity dependencies inherent in the knowledge graph. Nie et al. [31] proposed a context-based cross-language knowledge graph entity alignment method. This method generates entity and relationship features through embedding representation and Bi-LSTM model, and uses the graph attention mechanism to model neighbor categories and assign weights, thereby aligning entities within a unified vector space. Che et al. [5] proposed an improved cross-language entity alignment method. This method uses a bidirectional alignment graph convolutional network model that fuses attribute information, and combines feedforward neural network encoding entities with initial entity embedding to achieve cross-language entity alignment.

Cross-lingual entity alignment methods that utilize additional attributes, structures, and other information have demonstrated significant efficacy. However, these methods are heavily reliant on datasets with extensive human annotation and training, thereby escalating the manpower requirement. Consequently, the primary challenge addressed in this paper is to enhance the precision of entity alignment while minimizing the need for human intervention. Leveraging the superior comprehension and generalization capabilities of LLMs, we aim to efficiently execute the bilingual entity alignment task through model fine-tuning techniques. This approach seeks to optimize the entire process and elevate the alignment accuracy.

## 3. Materials and methods

### 3.1. Data Source.

PubMed is one of the most commonly used literature resources. This study used PubMed as the English data source and searched PubMed using the search term “coronavirus”, searching from December 2019 to June 2023. After manually excluding some irrelevant articles, a total of 18,687 articles were obtained.

We used 3 authoritative websites as Chinese data sources. UpToDate is a clinical decision support system based on the principles of evidence-based medicine [19]. UpToDate continuously combines the existing medical evidence with the clinical experience of experts to give a high level of practical medical information. UpToDate COVID-19 topic contains a variety of data resources such as clinical characteristics, diagnosis, patient management, and prevention of coronavirus.

National Science and Technology Library (NSTL) is a service system of scientific and technological literature information resources based on network environment [17]. NSTL collects and develops scientific and technological literature resources in various disciplines. The latest research progress of coronavirus was recorded under the topic of “Emerging Infectious Diseases” of NSTL.

SinoMed is a comprehensive biomedical literature service system, which integrates multiple resources such as China Biology Medicine disc (CBM) and Western Biology Medicine disc (WBM) [15].

A total of 17,653 coronavirus-related articles were retrieved from these three data sources, as shown in figure 1.

### 3.2. Knowledge Extraction.

SemRep is an effective off-the-shelf tool developed by the U.S. National Library of Medicine for entity and relationship identification [13]. The identified information can be linked to Unified Medical Language System(UMLS) standard terminology. We used SemRep to automatically extract English triples. Export the Titles and Abstracts information of the article in “PubMed” format. SemRep identifies entity and relation in each sentence and form into triples, then maps the entities to concept unique identifiers (CUIs) in the UMLS Metathesaurus, and the relationships map to UMLS Semantic Network. For Chinese data

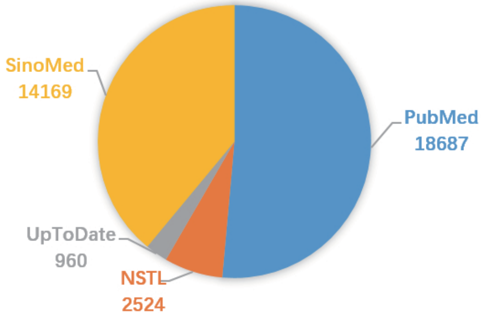

FIGURE 1. Statistics of Chinese and English data sources

sources, we used W2NER [14] model to extract entities and PL-Marker [27] model to extract relations. The Encoder Layer of W2NER framework was used to obtain the context word representation of the input sentence. The Convolution Layer builds the word matrix. The bilinear Layer of the Co-Predictor layer performs inference. PL-Marker was used to obtain the Solid Marker before and after subject and the Levitated Marker of object, and the feature expressions of multiple relation pairs were obtained for relation classification.

## 3.3. **Knowledge fusion.**

3.3.1. *Ontology mapping.* This study focuses on semantic types closely related to coronavirus, such as causative viruses, genes, drugs, examine, etc. Therefore, we screen UMLS Semantic Network, and the main types and relationships of attention are shown in Figure 2. Mapping Chinese-English entity types and establishing 7 "identical" relationships for the next step of entity fusion.

3.3.2. *Entity fusion based on LLMs.* LLMs is usually exposed to data in multiple languages during the pre-training phase, which gives them a certain cross-language processing capability. In translation tasks, LLMs can make full use of this cross-linguistic information to improve the accuracy and fluency of translation. In addition, LLMs can quickly adapt to new languages through fine-tuning or migration learning. In the bilingual entity fusion stage, we use LLM's rich semantic knowledge and powerful reasoning capabilities to promote the integration of bilingual knowledge.

Using LLMs for translation, and Chinese Unified Medical Language System (CUMLS) terms constructed by Institute of Medical Information Chinese Academy of Medical Sciences were selected for model fine-tuning. Calculate the similarity between the translated entities and the extracted Chinese data. If the similarity is greater than the threshold we defined, it means that the two are to be fused. Then it is judged whether the entity types are the same. If the entities are of the same type, the data will be fused and the entities will be stored in the database. If the entity types are different, a new entity will be added, the process is shown in Figure 3.

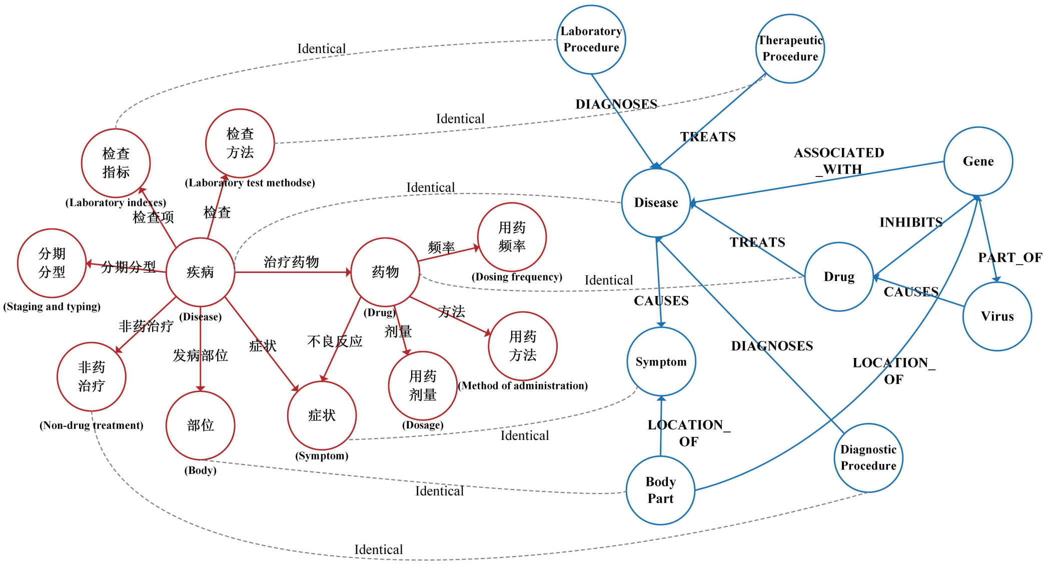

FIGURE 2. The main schema of the knowledge graph with the integration of bilingual biomedical knowledge

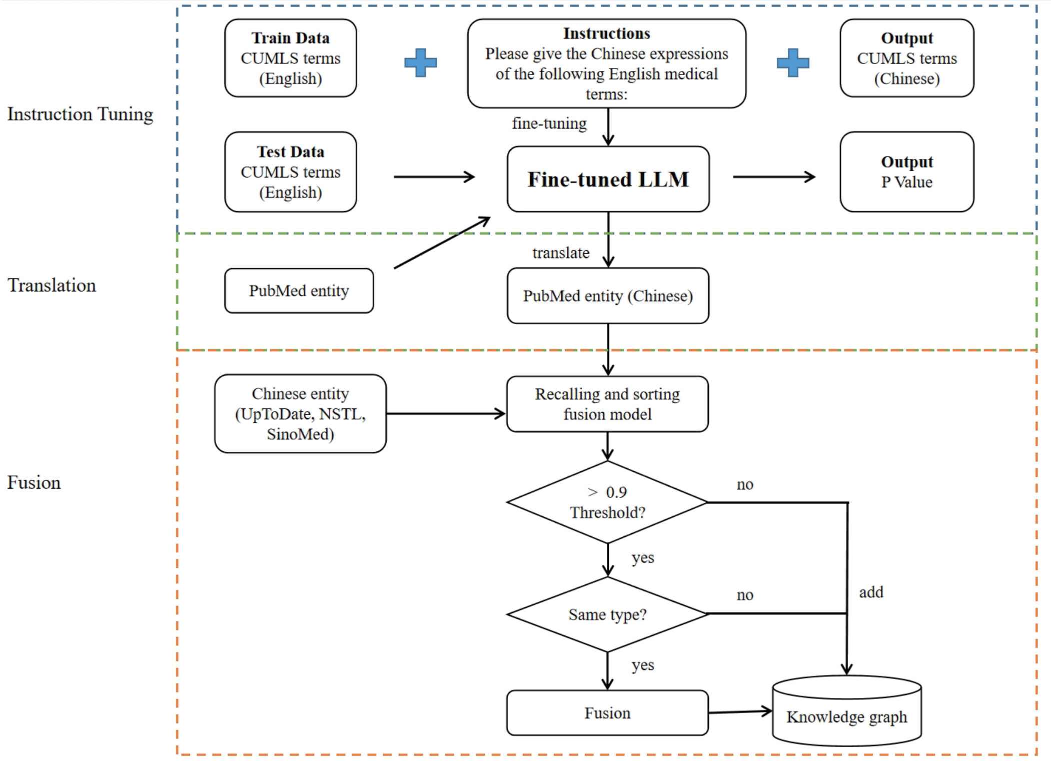

FIGURE 3. Diagram illustrating the workflow of our approach

The fine-tuning process aims to help LLMs better understand bilingual knowledge and further improve its accuracy and adaptability in fusion tasks. We constructed an instruction set for medical bilingual knowledge translation, including input, instructions, and output. The input is the English term in the CUMLS, the instructions is the task description, and the output is the corresponding Chinese term in the CUMLS. Based on the constructed instruction dataset, we perform instruction tuning on LLMs. Specifically, we choose the open source model Llama2-Chinese-13b-Chat [25] as our base LLMs and adopt the Lora method for fine-tuning.

Since the matching objects are massive entities and the complexity of fusion model is high, we propose a method of recalling entities first and then sorting them during similarity calculation. The recall phase gives priority to ensuring high recall rate and low time complexity. The sorting phase uses a higher accuracy algorithm to calculate the similarity between each entity. The similarity algorithm in the sorting stage is based on the recall algorithm and adds two data features for optimization.

In the recall phase, for a given entity $x$ and candidate entity set $\mathcal{Y}$, a method is designed to quickly find the top-$n$ candidate entity sets $\mathcal{Y} = \{y_1, y_2, \ldots, y_n\}$ with the highest similarity to $x$ from $\mathcal{Y}$. First, perform data cleaning. Data cleaning is used to remove symbols from entities and unify expressions such as "diabetes type 1" and "diabetes type I".

Then perform word segmentation. The Chinese entities are segmented by characters. For example, "xin xing guan zhuang bing du fei yan(COVID-19)" is split into "xin", "xing", "guan", "zhuang", "bing", "du", "fei", "yan". Finally, the Jaccard similarity algorithm is used to calculate the similarity between entity characters $x$ and $y_n$.

In the sorting phase, by observing the data, we found that medical entities can be split into two parts: subject and suffix. For example, "hu xi jiong po zong he zheng (respiratory distress syndrome)"can be split into "hu xi jiong po (respiratory distress)"and "zong he zheng (syndrome)", and "ao si ta wei jiao nang (oseltamivir capsules)"can be split into "ao si ta wei (oseltamivir)"and "jiao nang (capsules)". When the subject words are consistent and the suffixes are different, they often refer to the same or similar meaning; for subject words that are inconsistent, regardless of whether the suffixes are consistent or not, the similarity is low, such as "a mo xi lin jiao nang (amoxicillin capsule)"and "a si pi lin jiao nang (aspirin capsule)". By parsing the entity set, we construct a suffix vocabulary list. For entities $x$ and candidate entities $y_n$, calculate the subject word similarity and suffix similarity respectively. Secondly, we analyze the differences in the subject words, "fei chuan ran xing ji bing (non-infectious diseases)"and "chuan ran xing ji bing (infectious diseases)", the difference words is "fei (non)". The negative word has a more important influence on the meaning of the entity. When a negative word was present in the entity, two entities were considered not similar.

## 4. Results

4.1. **Data Extraction.** The Titles and Abstracts of the articles were downloaded in PubMed, and SemRep was used to extract coronavirus-related semantic information. Finally, 641,195 triples were obtained, which contained 13,065 concepts, 209 semantic types and 97 semantic relations in UMLS.

Extracting 17,653 coronavirus articles from 3 Chinese data sources: UpToDate, NSTL, and SinoMed. 32,432 triples were obtained, covering 7 entity types, 3 types of attributes, and 11 relation types.

4.2. **Data Fusion.** Dividing the CUMLS data into training and test sets in the ratio of 7:3. The main parameters were set as lora_rank 8, learning_rate 5e-5, learning_rate 5e-5, per_device_train_batch_size 8 for model fine-tuning. The translation accuracy of the fine-tuned LLMs on the test set was 73.72%. We used open source translation software to translate the test data, and the accuracy was 45.78%. The performance of LLMs on translation is much higher than that of existing tools. In the fusion stage, we set the threshold to 0.9 for data fusion, and 4330 entities can be fused, as detailed in Table 1.

Storing the coronavirus bilingual knowledge in a graph database, we use JavaScript to develop a B/S-based knowledge graph system, and various kinds of information such as pictures were added for multimodal display. The system embeds graph computing methods such as clustering and path calculation to facilitate knowledge discovery.

## 5. Discussion

By fusing the Chinese and English data, we have constructed a bilingual coronavirus knowledge graph with more comprehensive data, which can be used for broader knowledge discovery. The knowledge graph was used for drug and target knowledge discovery.

5.1. **Potential Drug.** By performing a search of the bilingual coronavirus knowledge graph, we found a “TREATS” relationship for COVID-19 for 5 drugs currently used to treat human immunodeficiency disease (HIV), as shown in Figure 4. According to the Anatomical Therapeutic Chemical (ATC) classification system of drugs, lopinavir, arbidol, and ribavirin are antiviral drugs for systemic use. The combination of lopinavir and ritonavir is often used to treat HIV-infected patients [16]. Studies have found that lopinavir can interact with the 3C-like chymotrypsin of coronavirus. In a clinical trial, it was found that lopinavir was more effective in treating hospitalized patients with severe COVID-19 than the control group [4]. Arbidol has strong inhibitory effect on a variety of viruses. Some researchers found that arbidol”s inhibitory efficiency against the new coronavirus at a concentration of $10 \sim 30\,\mu\mathrm{mol} \cdot L^{-1}$ was 60 times higher than that of the control group [1]. Arbidol can prevent the virus shell from contacting, adhering to, and fusion with the cell membrane of the host cell, thereby inhibiting the virus from entering the cell. Ribavirin is an antiretroviral drug. The drug enters cells and is phosphorylated to competitively inhibit the synthesis of viral guanosine triphosphate, thereby inhibiting the synthesis of viral mRNA [24]. In the “COVID-19 Diagnosis and Treatment Plan (Trial Sixth Edition)”, ribavirin is recommended to treat patients with COVID-19.

5.2. **Potential Target.** Searching for “coronavirus infection”, we obtain entities related to the coronavirus infection mechanism, as shown in Figure 5. Studies have found that angiotensin-converting enzyme 2(ACE 2) is a functional receptor for SARS-CoV-2 [3]. The recognition of viral proteins with cell surface ACE2 receptors

Table 1. Data fusion results

| Entity Type | Count | Fusion result examples |
|---|---|---|
| Disease | 1364 | {CH: 中东呼吸综合征;EN:Middle East respiratory syndrome} {CH: 结核病;EN:tuberculosis} {CH: 下呼吸道感染;EN:Lower respiratory tract infection} {CH: 气胸;EN:pneumothorax} {CH: 严重急性呼吸综合征;EN:Severe acute respiratory Syndrome} |
| Symptom | 877 | {CH: 咳嗽;EN:cough} {CH: 头晕;EN:dizzy} {CH: 咽痛;EN:pharyngalgia} {CH: 心动过缓;EN:bradycardia} {CH: 蛋白尿;EN:proteinuria} |
| Drug | 194 | {CH: 利托那韦;EN:Ritonavir} {CH: 伊马替尼;EN:Imatinib} {CH: 阿司匹林;EN:aspirin} {CH: 奥司他韦;EN:Oseltamivir} {CH: 利多卡因;EN:Lidocaine} |
| Laboratory Procedure | 996 | {CH: 中性粒细胞;EN:neutrophile granulocyte} {CH: 病毒核酸;EN:viral nucleic acid} {CH: 乳酸脱氢酶;EN:lactic dehydrogenase} {CH: 白细胞介素-4;EN:Interleukin-4} {CH: 免疫球蛋白G;EN:ImmunoglobulinG} |
| Body | 548 | {CH: 鼻窦;EN:paranasal sinus} {CH: 心内膜;EN:endocardium} {CH: 淋巴结;EN:lymph gland} {CH: 肺动脉;EN:pulmonary artery} {CH: 肾脏;EN:kidney} |
| Diagnostic Procedure | 201 | {CH: 情感支持;EN:emotional support} {CH: 无创通气;EN:noninvasive ventilation} {CH: 营养支持;EN:nutrition support} {CH: 血液透析;EN:hematodialysis} {CH: 免疫疗法;EN:immunotherapy} |
| Therapeutic Procedure | 150 | {CH: 体格检查;EN:physical examination} {CH: 超声检查;EN:supersonic inspection} {CH: 鼻拭子;EN:nose swab} {CH: 磁共振;EN:magnetic resonance} {CH: 肝功能检查;EN:liver function test} |

is the first step for viruses to infect host cells. Transmembrane protease serine 2 (TMPRSS2) on the cell surface initiate the activation of the spike protein, which then binds the spike protein to ACE2 and enters the host cell [6]. Although SARS-CoV-2 uses ACE2 as a receptor, studies have shown that the expression level of ACE2 is very low in most organs. Through gene expression similarity calculation, it was found that dipeptidyl peptidase 4 (DPP4) may be the virus of co-receptors. Several studies have predicted that DDP4 may be a candidate target of SARS-CoV-2 and participate in the coronavirus infection process [21].

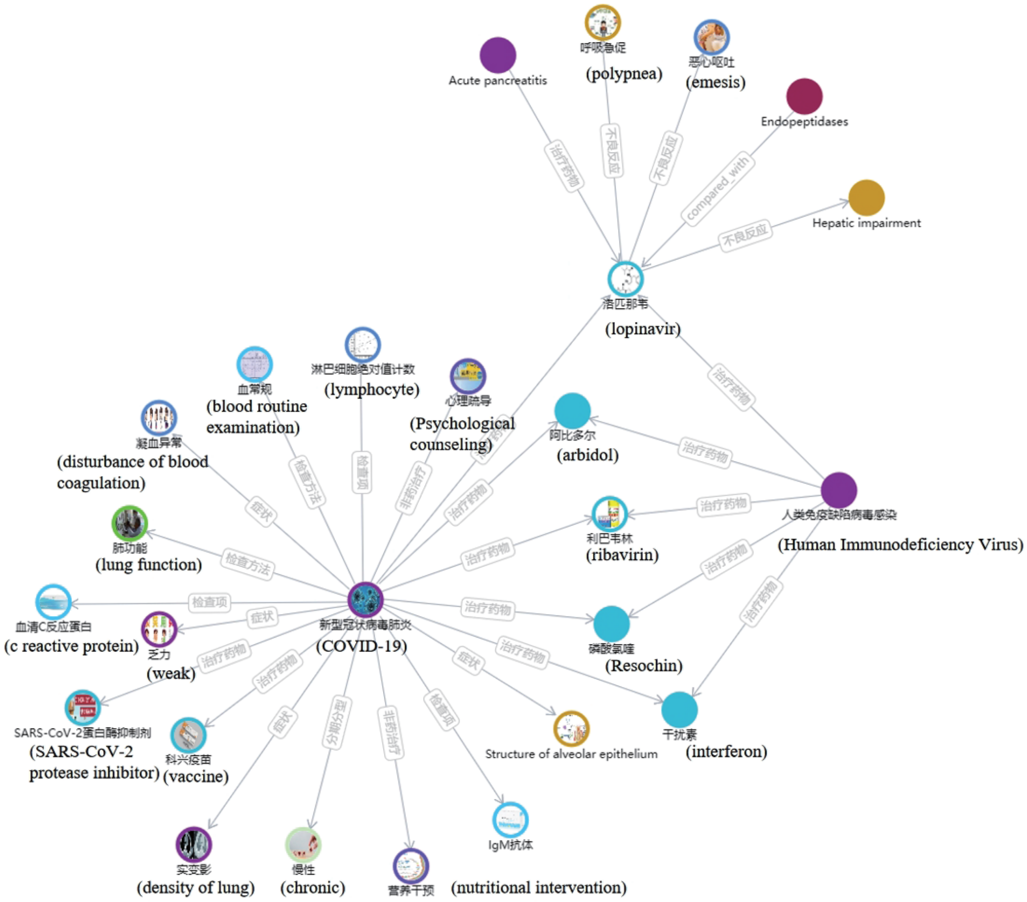

Figure 4. Drug repurposing for coronaviruses

## 6. Conclusions

In the context of globalization, the epidemic prevention, control and research of coronavirus require close cooperation among countries. However, language differences are a big barrier to cooperation. Consequently, this research introduces an LLMs-based cross-lingual data fusion approach, aiming to establish a method that is both cost-effective and highly accurate. The constructed bilingual knowledge graph of coronavirus effectively integrates heterogeneous information from multiple sources. The knowledge graph shows the characteristics of the virus, transmission mode, prevention measures and other key contents in a structured form, providing a comprehensive and easy-to-understand knowledge system for researchers and the public. In addition, the knowledge graph integrates information resources in multiple languages, so that researchers and the public from different language backgrounds can easily access and understand relevant information, and promote data sharing and cooperation on a global scale.

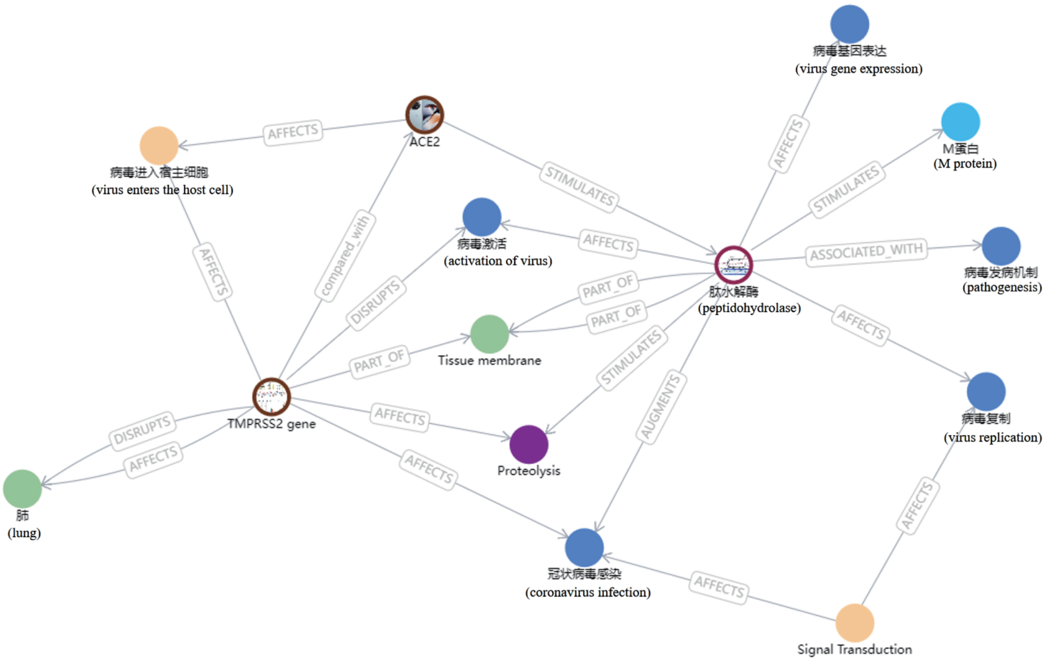

Figure 5. Potential target discovery

## References

[1] B. Amani, B. Amani, S. Zareei and M. Zareei, *Efficacy and safety of arbidol (umifenovir) in patients with COVID-19: A systematic review and meta-analysis*, Immun Inflamm Dis **9** (2021), 1197–1208.

[2] L. Bai, N. Li, G. Li, Z. Zhang and L. Zhu, *Embedding-based entity alignment of cross-lingual temporal knowledge graphs*, Neural Networks **172** (2024): 106143.

[3] S. Beyerstedt, E. B. Casaro and É. B. Rangel, *COVID-19: angiotensin-converting enzyme 2 (ACE2) expression and tissue susceptibility to SARS-CoV-2 infection*, European Journal of Clinical Microbiology & Infectious Diseases **40** (2021), 905–919.

[4] M. Biswas, *Predictive association of ABCB1 C3435T genetic polymorphism with the efficacy or safety of lopinavir and ritonavir in COVID-19 patients*, Pharmacogenomics **22** (2021), 375–381.

[5] C. Che and D. Liu, *Cross-language entity alignment based on bidirectional alignment and attribute information*, Computer Engineering **48** (2022), 74–80.

[6] I. J. Dos Santos Nascimento, E. F. da Silva-Júnior and T. M. de Aquino, *Molecular Modeling Targeting Transmembrane Serine Protease 2 (TMPRSS2) as an Alternative Drug Target Against Coronaviruses*, Curr Drug Targets **23** (2022), 240–259.

[7] F. Eggmann, R. Weiger, N. U. Zitzmann and M. B. Blatz, *Implications of large language models such as ChatGPT for dental medicine*, J. Esthet. Restor. Dent. **35** (2023), 1098–1102.

[8] M. Franco-Salvador, P. Gupta, P. Rosso and R. E. Banchs, *Cross-language plagiarism detection over continuous-space-and knowledge graph-based representations of language*, Knowledge-Based Systems **111** (2016), 87–99.

[9] B. Fu, R. Brennan and D. O'sullivan, *Cross-lingual ontology mapping–an investigation of the impact of machine translation*, in: Proceedings of Asian Semantic Web Conference, 2009, pp. 1–15.

[10] B. Fu, R. Brennan and D. O'sullivan, *Cross-lingual ontology mapping and its use on the multilingual semantic web*, in: the 19th International World Wide Web Conference (WWW 2010), Raleigh, USA, April 27th, 2010, P. Buitelaar, P. Cimiano, E. Montiel-Ponsoda, CEUR vol.571, 2010, pp. 13–20.

[11] K. Habas, C. Nganwuchu, F. Shahzad, R. Gopalan, M. Haque, S. Rahman, A. A. Majumder and T. Nasim, *Resolution of coronavirus disease 2019 (COVID-19)*, Expert Review of Anti-infective Therapy **18** (2020), 1201–1211.

[12] S. Kang, L. Ji, Z. Li, X. Hao and Y. Ding, *Iterative cross-lingual entity alignment based on TransC*, IEICE Trans. Inf. Syst **103-D** (2020), 1002–1005.

[13] H. Kilicoglu, G. Rosemblat, M. Fiszman and D. Shin, *Broad-coverage biomedical relation extraction with SemRep*, BMC Bioinformatics **21** (2020): 188.

[14] J. Li, H. Fei, J. Liu, S. Wu, M. Zhang, C. Teng, D. Ji, F Li, *Unified named entity recognition as word-word relation classification*, ArXiv, (2021), abs.2112.10070.

[15] K. Liu, W. Zhang, W. Li, T. Wang and Y. Zheng, *Effectiveness of virtual reality in nursing education: a systematic review and meta-analysis*, BMC Medical Education **23** (2023): 710.

[16] P. Magro, I. Zanella, M. Pescarolo, F. Castelli and E. Quiros-Roldan, *Lopinavir/ritonavir: repurposing an old drug for HIV infection in COVID-19 treatment*, Biomedical Journal **44** (2021), 43–53.

[17] L. S. Meng, *On the development of national science and technology library*, Interlending & Document Supply **42** (2014), 131–136.

[18] M. Roostaee, S. M. Fakhrahmad and M. H. Sadreddini, *Cross-language text alignment: A proposed two-level matching scheme for plagiarism detection*, Expert Systems with Applications **160** (2020): 113718.

[19] L. S. Ensan, M. Faghankhani, A. Javanbakht, S.-F. Ahmadi and H. R. Baradaran, *To compare PubMed clinical queries and UpToDate in teaching information mastery to clinical residents: A crossover randomized controlled trial*, Plos One **6** (2011): e23487.

[20] N. H. Shah, D. Entwistle and M. A. Pfeffer, *Creation and adoption of large language models in medicine*, Journal of the American Medical Association **330** (2023), 866–869.

[21] S. B. Solerte, A. Di Sabatino, M. Galli and P. Fiorina, *Dipeptidyl peptidase-4 (DPP4) inhibition in COVID-19*, Acta Diabetologica **57** (2020), 779–783.

[22] D. Spohr, L. Hollink and P Cimiano, *A machine learning approach to multilingual and cross-lingual ontology matching*,in: Proceedings of International Semantic Web Conference 2011, pp. 665–680.

[23] M. Sreepadmanabh, A. K. Sahu and A. Chande, *COVID-19: Advances in diagnostic tools, treatment strategies, and vaccine development*, Journal of Biosciences **45** (2020): 148.

[24] S. Tejada, R. Martinez-Reviejo, H. N. Karakoc, Y. Peña-López, O. Manuel and J. Rello, *Ribavirin for treatment of subjects with respiratory syncytial virus-related infection: A systematic review and meta-analysis*, Adv Ther **39** (2022), 4037–4051.

[25] H. Touvron, L. Martin, K. R. Stone, P. Albert, A. Almahairi, Y. Babaei and N. Bashlykov, *Llama 2: Open foundation and fine-tuned chat models*, ArXiv (2023), abs/2307.09288.

[26] C. X. Wang, Z. H. Huang, Y. Wan, J. Y. Wei, J. Z. Zhao and P. H. Wang, *FuAlign: Cross-lingual entity alignment via multi-view representation learning of fused knowledge graphs*, Inf. Fusion **89** (2022), 41-52.

[27] D. M. Ye, Y. K. Lin, P. Li and M. S. Sun, *Packed levitated marker for entity and relation extraction*, ArXiv (2021), abs.2109.06067.

[28] Z. Ye, J. Huang, B. He and H. Lin, *Mining a multilingual association dictionary from Wikipedia for cross-language information retrieval*, Journal of the Association for Information Science and Technology **63** (2012), 2474–2487.

[29] X. Zhang, W. Zhang and H. Wang, *Cross-language entity alignment based on dual-relation graph and neighbor entity screening*, Electronics **12** (2023): 1211.

[30] Z. Zhao and S. Lin, *A cross-linguistic entity alignment method based on graph convolutional neural network and graph attention network*, Computing **105** (2023), 2293–2310.

[31] B. B. Zhu, T. Bao, L. Liu, J.Y. Han, J.Y. Wang and T. Peng, *Cross-lingual knowledge graph entity alignment based on relation awareness and attribute involvement*, Applied Intelligence **53** (2022), 6159–6177.

*Manuscript received March 25, 2024*
*revised August 29, 2024*

Pei Lou
Institute of Medical Information, Chinese Academy of Medical Sciences & Peking Union Medical College, Peking, China
*E-mail address*: 307092141@qq.com

An Fang
Institute of Medical Information, Chinese Academy of Medical Sciences & Peking Union Medical College, Peking, China
*E-mail address*: fang.an@imicams.ac.cn

Kuanda Yao
Institute of Medical Information, Chinese Academy of Medical Sciences & Peking Union Medical College, Peking, China
*E-mail address*: yao.kuanda@imicams.ac.cn

Wanqing Zhao
Institute of Medical Information, Chinese Academy of Medical Sciences & Peking Union Medical College, Peking, China
*E-mail address*: zhao.wanqing@imicams.ac.cn

Chenliu Yang
Institute of Medical Information, Chinese Academy of Medical Sciences & Peking Union Medical College, Peking, China
*E-mail address*: yang.chenliu@imicams.ac.cn

Jiahui Hu
Institute of Medical Information, Chinese Academy of Medical Sciences & Peking Union Medical College, Peking, China
*E-mail address*: hu.jiahui@imicams.ac.cn