Computers, Materials & Continua

Tech Science Press

DOI: 10.32604/cmc.2024.053506

ARTICLE

GATiT: An Intelligent Diagnosis Model Based on Graph Attention Network Incorporating Text Representation in Knowledge Reasoning

Yu Song, Pengcheng Wu, Dongming Dai, Mingyu Gui and Kunli Zhang*

School of Computer and Artificial Intelligence, Zhengzhou University, Zhengzhou, 450001, China

*Corresponding Author: Kunli Zhang. Email: ieklzhang@zzu.edu.cn

Received: 02 May 2024 Accepted: 16 August 2024 Published: 12 September 2024

ABSTRACT

The growing prevalence of knowledge reasoning using knowledge graphs (KGs) has substantially improved the accuracy and efficiency of intelligent medical diagnosis. However, current models primarily integrate electronic medical records (EMRs) and KGs into the knowledge reasoning process, ignoring the differing significance of various types of knowledge in EMRs and the diverse data types present in the text. To better integrate EMR text information, we propose a novel intelligent diagnostic model named the Graph ATtention network incorporating Text representation in knowledge reasoning (GATiT), which comprises text representation, subgraph construction, knowledge reasoning, and diagnostic classification. In the text representation process, GATiT uses a pre-trained model to obtain text representations of the EMRs and additionally enhances embeddings by including chief complaint information and numerical information in the input. In the subgraph construction process, GATiT constructs text subgraphs and disease subgraphs from the KG, utilizing EMR text and the disease to be diagnosed. To differentiate the varying importance of nodes within the subgraphs features such as node categories, relevance scores, and other relevant factors are introduced into the text subgraph. The message-passing strategy and attention weight calculation of the graph attention network are adjusted to learn these features in the knowledge reasoning process. Finally, in the diagnostic classification process, the interactive attention-based fusion method integrates the results of knowledge reasoning with text representations to produce the final diagnosis results. Experimental results on multi-label and single-label EMR datasets demonstrate the model's superiority over several state-of-the-art methods.

KEYWORDS

Intelligent diagnosis; knowledge graph; graph attention network; knowledge reasoning

1 Introduction

Recently, the medical field has been exploring the application of artificial intelligence to improve the quality and efficiency of medical services [1]. Electronic medical records (EMRs) encompass structured information detailing the patient's diagnosis and treatment process, as well as unstructured textual data representing doctor diagnoses and treatment opinions. A physician's clinical diagnostic process involves evaluating the likelihood of a patient having a particular disease based on their clinical manifestations and examination results. Typically, EMRs contain one or more diagnoses. For instance, a patient may receive diagnoses for both "diabetes" and "hypertension". Viewing an EMR as a sample

can pertain to multiple classifications. Thus, the problem of intelligent diagnosis can be considered a classification task, with each diagnosis in a medical record representing a distinct label [2–4].

Diagnosing and treating a patient's disease involves using both medical knowledge and the patient's EMR. Knowledge graphs (KGs) have emerged as a critical knowledge source in various knowledge-driven algorithms and systems [5]. KGs consist of entities and the relations between them, denoted as $\mathcal{G} = (\mathcal{V}, \mathcal{E})$, where $\mathcal{V}$ and $\mathcal{E}$ denote all entities and relations in the graph [6]. Integrating medical KGs into intelligent diagnostic tasks enriches diagnostic models with specialized knowledge, thereby enhancing their performance. The KG contains a wealth of valuable implicit knowledge, which cannot be fully extracted through distribution-based knowledge reasoning alone. Neural network-based knowledge reasoning uncovers hidden knowledge within the graph, enhancing reasoning capabilities [7]. Notably, knowledge reasoning based on graph neural networks (GNNs) [8] can adaptively learn features and topologies within the graph in an end-to-end manner. Sun et al. [9] leveraged graph structure information to uncover complex nonlinear relations, thereby improving the accuracy and efficiency of knowledge reasoning.

Currently, the challenges of EMR-based intelligent diagnostic models are as follows:

- Current research addressing EMRs through direct encoding for text representation often ignores the heterogeneity of information across different sections of EMRs. The chief complaint information in EMRs is more important for diagnostic output compared to plain text information. Additionally, numerical information cannot be directly represented by the model. How to represent the different types of text information in EMRs is a challenge.
- Existing models improve diagnosis by incorporating external KGs introduced based on the diseases to be diagnosed, failing to notice critical information in EMRs. Furthermore, the distributed-based knowledge reasoning method is unable to effectively extract the hidden features and topologies within the KG. Therefore, it is a challenge to effectively utilize critical text information in EMRs and how to better learn implicit information about KG for diagnosis during the knowledge reasoning process.

In light of the aforementioned challenges, we consider the varying impact on diagnostic output and the diverse forms of expression. Consequently, information within EMRs can be classified into three categories: ordinary text, chief complaints, and numerical information. The chief complaint, which represents the primary pain or most apparent symptom experienced by the patient, has significance in the diagnostic process that outweighs other text information. Numerical information refers to specific test results and clinical signs recorded as numeric values in EMRs. MacBERT (MLM as correction, Mac) [10] improves the original BERT (Bidirectional Encoder Representations from Transformers) [11] model through the use of MLM (Masked Language Modeling) as a refined correction technique, optimizing its performance for more accurate natural language understanding. To address the first challenge, we incorporated the chief complaint information into the MacBERT [10] encoding and combined it with the numerical data to derive the textual representation of EMRs.

Furthermore, we construct disease subgraphs and text subgraphs to integrate EMR representation into knowledge reasoning. Disease subgraphs are identified from the KG based on the list of diseases to be diagnosed. Text subgraphs are obtained based on the entities in the EMR after retrieving them from the KG. To address the second challenge, we incorporate context nodes containing EMR information into the knowledge reasoning process and set the node category to distinguish them from the rest of the entities. The information contained within context nodes differs from that in other nodes, and the significance of entities directly extracted from EMR text varies compared to other entities. Fig. 1 illustrates a text subgraph constructed from the EMR text "In the past month, multiple

measurements of elevated blood pressure, ..., worsening dizziness without apparent cause, ...". For instance, if the final diagnosis is "hypertension", the text subgraph should contain two entity nodes: "elevated blood pressure" and "dizziness". Notably, the former is significantly more crucial to the final diagnosis than the latter, which is reflected in the node relevance score obtained by the model. The Graph Attention Network (GAT) [12] employed in this study utilizes an attention mechanism to assign different weights according to the relative importance among nodes, implementing various learning strategies to effectively address node disparities.

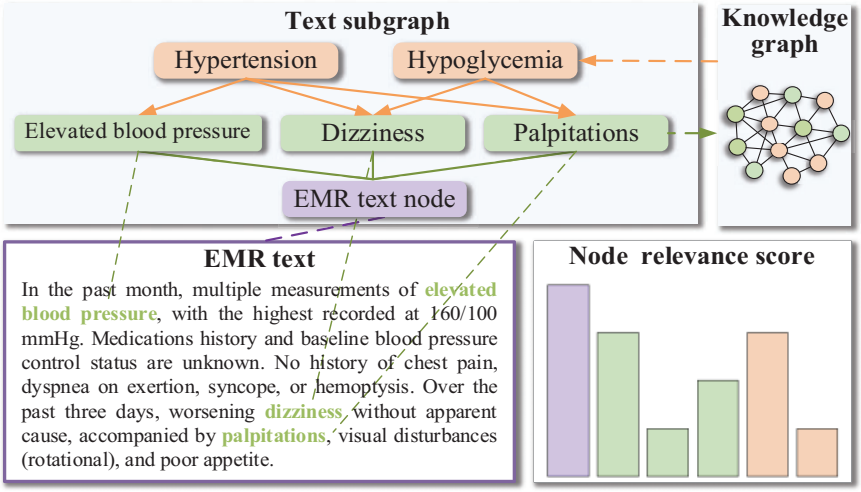

Figure 1: A sample of constructing a text subgraph

Altogether, the following are the primary contributions of our work:

- We propose GATiT, an intelligent diagnosis model based on GAT. We constructed disease subgraphs and text subgraphs, integrating disease information and EMR text information in the knowledge reasoning process. By incorporating node categories, node relevance scores, and edge relevance scores, we optimize GAT for knowledge reasoning. GAT can assign varying weights according to the relative importance of nodes and can better address the differences between nodes.
- GATiT encodes text representations of EMRs using MacBERT, which incorporates enhancements for chief complaint and numerical information. We introduce *ChiefInfo Embedding*, which integrates chief complaint information into the inputs of MacBERT and utilizes self-attention to fuse numerical and textual information, thereby facilitating the learning of various categories of information from EMRs.
- Comparative experiments conducted on multi-label and single-label EMR datasets using GATiT demonstrate its superiority over several state-of-the-art methods, as indicated by the analysis of experimental results.

2 Related Work

Many researchers have studied intelligent diagnosis methods extensively. In this section, in addition to describing existing intelligent diagnosis algorithms (Section 2.1), recent work on graph neural networks is presented (Section 2.2).

2.1 Intelligent Diagnosis

Intelligent methods for disease diagnosis have received a lot of attention in recent years. Li et al. [13] proposed a deep learning framework and fine-tuned a CNN (Convolutional Neural Networks) model for intelligent-assisted diagnosis in pediatrics. The results on real-world pediatric Chinese EMR data demonstrated that this method achieved an average accuracy and F1-score of 81%. Sun et al. [14] explored a privacy-preserving medical record searching scheme (PMRSS) utilizing the ElGamal Blind Signature. This scheme allows patients to independently make medical diagnoses by securely searching and comparing previous and current medical records.

While the mentioned methods have been effective, they primarily focus on data processing and neglect the integration of medical domain knowledge crucial for physicians' diagnostic processes. With the advancement of medical KGs, there is a growing interest in intelligent diagnosis research that incorporates external medical knowledge. Chen et al. [15] introduced a Sequence-to-Subgraph framework and devised a subgraph convolutional network and hierarchical diagnostic attentive network model (SHiDAN) based on this framework. SHiDAN organizes EMR text into densely connected subgraphs with external medical knowledge, employing hierarchical feature extraction for patient diagnosis. Yang et al. [16] proposed a time-aware KG attention approach to address the problem of knowledge decay over time and used a comprehensive representation of the local KG to select candidate global knowledge for prospective interpretation.

2.2 Graph Neural Network

GNNs are extensively utilized in disease diagnosis. Yang et al. [17] proposed a feature aggregation-based intelligent diagnostic model using heterogeneous graph convolutional networks (GCNs) that emphasizes the intrinsic properties of symptoms and the multiple hidden relationships among them for effective and accurate symptom-based knowledge reasoning and disease diagnosis. Song et al. [18] incorporated disease subgraph embedding into the knowledge reasoning process using GCNs to amplify the relevance of disease-related knowledge, yet relying solely on entity subgraphs in the knowledge reasoning process presents limitations in effectively combining simulated medical knowledge with real patient cases.

According to the aforementioned studies related to deep learning-based and GNN-based intelligent diagnosis, most of the approaches have been able to obtain satisfactory results in disease diagnosis through deep learning models and disease enhancement. However, most existing intelligent diagnosis models enhance their performance by incorporating external KGs focused solely on the diseases listed in the EMR, neglecting the textual information within the EMRs. Meanwhile, most of the work introduces only EMR text and does not consider the heterogeneity of data in EMR, which leads to unsatisfactory results. Therefore, intelligent diagnosis study should focus on enhancing knowledge around EMR and external KG, and different categories of information from EMR text to obtain more accurate text representation.

3 Model

The architecture of our model is shown in Fig. 2. This section describes the text representation (Section 3.1), the construction process of both the text subgraph and the disease subgraph (Section 3.2), and the knowledge reasoning (Section 3.3), and the diagnostic classification (Section 3.4).

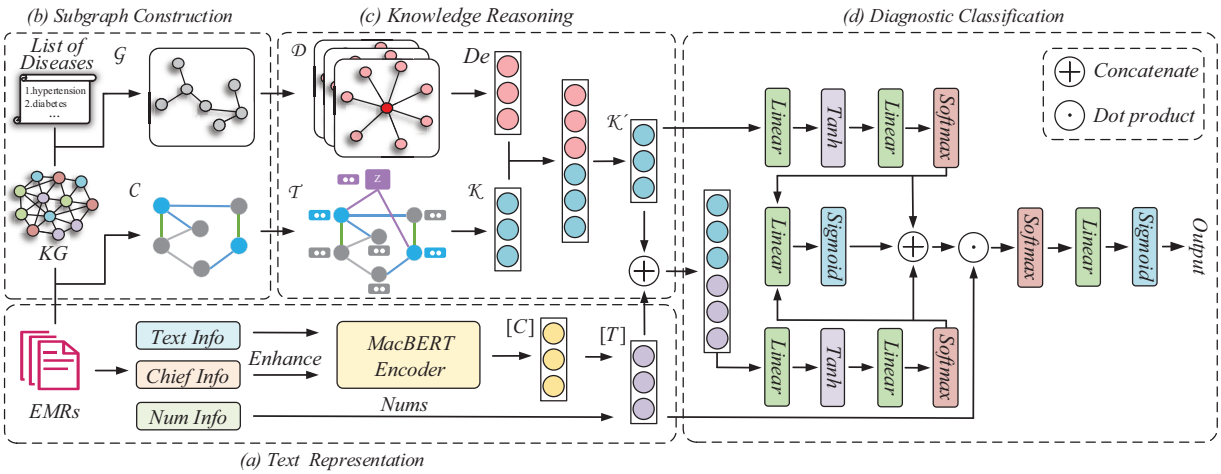

Figure 2: Illustration of GATiT. Our model is structured with text representation, subgraph construction, knowledge reasoning, and diagnostic classification. GATiT integrates knowledge from KG and text information from EMRs for intelligent diagnosis tasks. (a) In the text representation, *TextInfo*, *ChiefInfo*, and *NumInfo* from EMRs are encoded using MacBERT, incorporating chief complaint enhancement and numerical information injection to produce the final text representation. (b) In the subgraph construction, the disease subgraph $\mathcal{D}$ extracted from the list of diseases to be diagnosed and the text subgraph $\mathcal{T}$ extracted from the EMRs and KG. Incorporating the EMR context node z, node category features, node relevance scores, and edge additional features into $\mathcal{T}$ helps to distinguish nodes. (c) The message-passing strategy and attention weight calculation of GAT are adjusted to learn these features in the knowledge reasoning. (d) Finally, the text representations and knowledge reasoning outputs are fused using interactive attention mechanisms in the diagnostic classification

3.1 Text Representation

The text representation utilizes MacBERT and incorporates chief complaint enhancement (Section 3.1.1) and numerical information injection (Section 3.1.2) for encoding the text representation of EMRs. We define ordinary text information as ***TextInfo***, chief complaint information as ***ChiefInfo***, and numerical information as ***NumInfo***.

3.1.1 Chief Complaint Enhancement

The original input of MacBERT includes *Token Embedding*, *Segment Embedding*, and *Position Embedding*. Since the input to GATiT is a sequence, all values in the *Segment Embedding* are set to zero. *Position Embedding* is employed to retain information about the semantic order of the input sequence or sequence pair. Building upon the achievements of Qu et al. [19] in integrating history response embedding into a dialogue system, GATiT adds ***Chief Embedding*** containing the chief complaint

information to the MacBERT input. ***Chief Embedding*** labels the chief complaint information in the input sequence as E_{CC}, and the rest of the characters as E_G.

3.1.2 Numerical Information Injection

MacBERT encounters challenges in learning numerical features from EMRs. Therefore, we extract numerical features from the EMR, and the hidden layer representation [*C*], enhanced by chief complaint information, is concatenated with the numerical information and subsequently fused using a multi-head attention mechanism. The specific procedure is outlined as follows:

$$Q = K = V = Concat\ ([C];\ Nums) \tag{1}$$

$$Attention\ (Q,\ K,\ V) = softmax\ \left(\frac{QK^{\top}}{\sqrt{d_k}}\right) V \tag{2}$$

$$head_i = Attention\ (\mathbf{W}_i^Q Q,\ \mathbf{W}_i^K K,\ \mathbf{W}_i^V V) \tag{3}$$

$$[T] = \mathbf{W}^O Concat\ (head_0;\ head_1;\ \ldots;\ head_i) \tag{4}$$

where [*C*] represents the hidden layer representation marked in MacBERT after the enhancement with chief complaints information, which contains information on the entire input sequence. *Nums* represents the processed numerical features. $\mathbf{W}^Q$, $\mathbf{W}^K$, $\mathbf{W}^V$, and $\mathbf{W}^O$ represent trainable matrices. [*T*] represents the text representation after enhancing chief complaints and injecting numerical information.

3.2 Subgraph Construction

The process of the subgraph construction is depicted in Fig. 3. This section constructs the text subgraph (Section 3.2.1) and the disease subgraph (Section 3.2.2) within the EMR text, facilitating their application in the subsequent knowledge reasoning and establishing a close connection between the knowledge reasoning and EMR text information.

3.2.1 Text Subgraph

For a given EMR text sequence $\mathcal{S} = (s_0, s_1, \ldots, s_{n-1})$ and KG $\mathcal{G} = (\mathcal{V},\ \mathcal{E})$, retrieving entities corresponding to the $\mathcal{G}$ from the EMR text through the similarity algorithm and construct the corresponding entity subgraph $\mathcal{C}_i = (\mathcal{V}_{en},\ \mathcal{E}_{en})$, where $\mathcal{V}_{en}$ consists of entities identified in the text (disease and symptom), as well as tail entities or head entities corresponding to these entities retrieved from triples, and $\mathcal{E}_{en}$ represents the relations between all entities in $\mathcal{V}_{en}$ [18].

EMR context node *z*. In order to design a joint reasoning space for text representations and text subgraphs of EMRs, the model explicitly connects them in a common graph structure. z is defined within the set of nodes of the original entity subgraph, housing the text representation of EMRs. Notably, z is not a node in $\mathcal{G}$ and cannot be initialized using a distributed knowledge-based reasoning model. Therefore, the model initializes z using a projection matrix that projects z into the joint reasoning space.

$$z_{PTM} = W_L(f_{PTM}(text(z))) \tag{5}$$

where z_{PTM} represents the initialization vector of the text representation node z. Here, f_{PTM} signifies the pre-trained model, while W_L denotes the projection matrix. Notably, the remaining nodes in $\mathcal{C}$ are initialized using TransR [20].

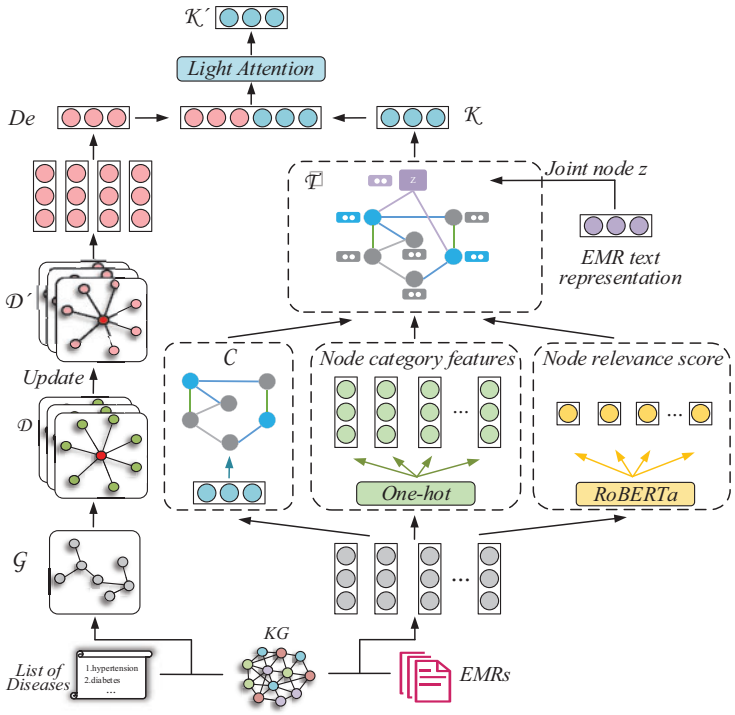

Figure 3: Illustration of the subgraph construction and the knowledge reasoning. The disease subgraphs $\mathcal{D}$ is obtained by retrieving triples from the KG based on the list of diseases to be diagnosed. The vector representation ***De*** is generated following batch updates. The entity subgraphs $\mathcal{C}$ are obtained by retrieving entities in the KG after they have been identified in the EMR. The text subgraphs $\mathcal{T}$ were created by incorporating the EMR context node (z), node category features, and node relevance scores within $\mathcal{C}$. $\mathcal{T}$ utilizes the message-passing mechanism and attention-weighting computation of GAT to learn features that differentiate the relevance of nodes for knowledge reasoning

Node category features. Following the construction process, nodes can be categorized into three types: (a) Context node, which contains the text representation of EMRs. (b) Text entity node, which corresponds to disease entities and symptom entities identified from the EMR text. (c) Other entity nodes, which represent the head entity or tail entity nodes in the triple retrieved from KGs by the text entity nodes. The category feature of a node is defined as:

$$Type = (Z,\ T,\ O) \tag{6}$$

where Z represents EMR text nodes, T represents text entity nodes, and O represents other entity nodes. A corresponding one-hot vector $u_i \in \{0,1\}^{|Types|}$ is defined for $v_i \in \mathcal{V}_{sub}$ to represent the node category features.

Node relevance scores. Nodes within the same category in $\mathcal{T}$ exhibit varying levels of importance towards the final diagnostic output. However, the construction process of $\mathcal{T}$ unavoidably introduces some noise nodes. A pre-trained model assigns a relevance score to each node to gauge their significance and facilitate node filtering. Specifically, for $v_i \in \mathcal{V}'$, the relevance score is computed by concatenating the $text(z)$ and the $text(v_i)$ associated with v_i:

$$\rho_i = f'_{PTM}(Concat(text(z); text(v_i))) \tag{7}$$

where f'_{PTM} represents RoBERTa (Robustly Optimized BERT Pretraining Approach) [21]. ρ_i quantifies the significance of each node within $\mathfrak{T}$ concerning its associated EMR text.

Edge additional features. Relation type r' describes the relation between z and the nodes of the text entity. Furthermore, the one-hot vector $e_{ij} \in \{0, 1\}^{|R|}$ is introduced to characterize the edge category corresponding to node v_i and the relation r_{ij} between nodes v_j. The number of edge categories $|R|$ is defined as:

$$|R| = 2 * (|r| + 1) \tag{8}$$

where $|r|$ represents KG's total number of relation categories. Specifically, 1 denotes the relation r' between z and the text entity. Since $\mathfrak{T}$ defined by the model is an undirected graph, each edge is duplicated to incorporate the corresponding reversed edges, thus doubling the edge count and defining the corresponding edge categories accordingly.

$\mathfrak{T}$ is defined as:

$$\mathfrak{T} = (\mathcal{V}_{sub}, \mathcal{E}_{sub}) \tag{9}$$

$$\mathcal{V}_{sub} = \mathcal{V}_{en} \bigcup \{z\} \tag{10}$$

$$\mathcal{E}_{sub} = \mathcal{E}_{en} \bigcup \{(z, r', v) | v \in \mathcal{V}_T\} \tag{11}$$

where z represents a context node containing a text representation of the EMR, $\mathcal{V}_T$ represents a collection of entities recognized directly from the text, and r' represents a relation between the EMR text node z and each entity recognized directly from the text.

3.2.2 Disease Subgraph

Incorporating disease subgraph ($\mathfrak{D}$) embeddings into the knowledge reasoning process enhances the understanding of the disease under-diagnosis, facilitating the model in establishing connections between $\mathfrak{T}$ and the target disease.

$\mathfrak{D}$ is extracted from the $\mathfrak{G}$ using the set of diseases to be diagnosed from the EMR dataset through the similarity algorithm. The construction process of $\mathfrak{D}$ is the same as $\mathfrak{C}$ [18].

$\mathfrak{D}$ consists of n connected subgraphs. Parameters are shared across the n connected subgraphs to reduce parameter requirements and accelerate training. These n connected subgraphs are grouped into a graph batch, with each connected subgraph corresponding to a distinct $\mathfrak{D}$. The connectivity subgraphs are updated dynamically as follows:

$$Dis_i^{(l+1)} = ReLU(b_1^{(l)} + \sum\nolimits_{j \in \mathfrak{N}(i)} \frac{h_j^{(l)} \mathbf{W}_1{}^{(l)}}{\sqrt{|\mathfrak{N}(i)|}} + \frac{Dis_i^{(l)} \mathbf{W}_2{}^{(l)}}{|\mathfrak{N}(i)|}) \tag{12}$$

$$h_j^{(l+1)} = ReLU(b_2^{(l)} + \frac{Dis_i^{(l)} \mathbf{W}_3{}^{(l)}}{\sqrt{|\mathfrak{N}(i)|}} + \mathbf{W}_4{}^{(l)} h_j^{(l)}) \tag{13}$$

where $\mathfrak{N}(i)$ represents the set of neighboring nodes of node i, $Dis_i^{(l)}$ represents the representation of the ith diseased node in layer l, $h_j^{(l)}$ represents the representation of the jth non-diseased node in layer l. $\mathbf{W}_1{}^{(l)}$, $\mathbf{W}_2{}^{(l)}$, $\mathbf{W}_3{}^{(l)}$, $\mathbf{W}_4{}^{(l)}$, $b_1^{(l)}$, $b_2^{(l)}$ are the trainable parameters in layer l.

Splitting the whole graph batch gives updated $\mathfrak{D}$:

$$\mathfrak{D}' = (\mathfrak{D}_1', \mathfrak{D}_2', \ldots, \mathfrak{D}_{n-1}') \tag{14}$$

The vector representation of the final subgraph is obtained by reading $\mathfrak{D}'$:

$$De = (e_0, e_1, \ldots, e_{n-1}) \tag{15}$$

where e_i represents the vector representation corresponding to the ith connected subgraph $\mathfrak{D}_i$.

3.3 Knowledge Reasoning

The architecture of knowledge reasoning is depicted in Fig. 3. Knowledge reasoning uses $\mathfrak{T}$ and $\mathfrak{D}$, which is described in Section 3.2. This section is categorized into messaging policy (Section 3.3.1), attention weighting (Section 3.3.2), and subgraph reasoning (Section 3.3.3).

3.3.1 Messaging Policy

Aggregation of messages between each node and neighboring nodes using an attention mechanism:

$$h_i^{(l+1)} = ReLU\left(\sum_{j \in N_i} \alpha_{ij} \mathbf{W_5} h_j^{(l)}\right) \tag{16}$$

where $h_i^{(l)}$ denotes the hidden representation of node v_i in layer l of the model, N_i represents the set of neighboring nodes of node v_i, α_{ij} signifies the attention coefficients from node v_i to node v_j, and $\mathbf{W_5}$ denotes the corresponding weight matrix. The attention coefficients α_{ij} are computed within GAT using a learnable function, which enables the model to detect distinctions between nodes.

Unlike GAT, the message m_{ij} transmitted from node v_i to node v_j includes not only the hidden layer representation of node v_i, but the category features of the relation r_{ij} between v_i and v_j, and of nodes v_i and v_j themselves. The definition of m_{ij} is as follows:

$$u_i' = f_u(u_i) \tag{17}$$

$$r_{ij} = f_r(Concat(e_{ij}; u_i; u_j)) \tag{18}$$

$$m_{ij} = f_m(Concat(h_i^l; u_i'; r_{ij})) \tag{19}$$

where $u_i, u_j \in \{0,1\}^{|\text{Types}|}$ is a one-hot vector representing the category features of nodes v_i and v_j, $e_{ij} \in \{0,1\}^{|R|}$ is a one-hot vector representing the category of the relation r_{ij}. h_i^l represents node v_i's hidden representation in the layer l model. Let the dimension to which the final reasoning space belongs be $\mathbb{R}^D$, f_u: $\mathbb{R}^{|\text{Types}|} \rightarrow \mathbb{R}^{D/2}$ is a projection of the node category vectors into the space of the $\mathbb{R}^{D/2}$ linear transformation layer, f_r: $\mathbb{R}^{2\times|\text{Types}|+|R|} \rightarrow \mathbb{R}^D$ is a multilayer perceptron (MLP) that aggregates the node v_i category features, the node v_j category features and the relation r_{ij} category features, f_m: $\mathbb{R}^{2.5D} \rightarrow \mathbb{R}^D$ is a linear transformation layer that projects all the information conveyed by node v_i to node v_j to the $\mathbb{R}^{D/2}$ space.

3.3.2 Attention Weighting

The attention coefficient α_{ij} between nodes v_i and v_j denotes the strength of their association, jointly determined by their types, the relation representation, and the relevance score. The calculation

process of α_{ij} is as follows:

$$\rho_i^{'} = f_\rho(\rho_i) \tag{20}$$

$$q_i = f_q(Concat\ (h_i^l;\ u_i^{'};\ \rho_i^{'})) \tag{21}$$

$$k_j = f_k(Concat(h_j^l; u_j^{'}; \rho_j^{'}; r_{ij})) \tag{22}$$

$$\gamma_{ij} = \frac{q_i^\top k_j}{\sqrt{D}} \tag{23}$$

$$\alpha_{ij} = \frac{\exp(\gamma_{ij})}{\sum_{j' \in N_i \cup \{i\}} \exp(\gamma_{ij'})} \tag{24}$$

where $f_\rho : \mathbb{R} \to \mathbb{R}^{D/2}$ is an MLP mapping node relevance scores into the space of $\mathbb{R}^{D/2}$, $f_q : \mathbb{R}^{2D} \to \mathbb{R}^D$ is a linear transformation layer mapping all the features of node v_i into the space of $\mathbb{R}^D$, q represents the query matrix summarized by the attention mechanism, and $f_k : \mathbb{R}^{3D} \to \mathbb{R}^D$ is an MLP mapping the information related to node v_j into the space $\mathbb{R}^D$ in a linear transformation layer, k represents the key matrix in the attention mechanism, and $\sqrt{D}$ represents the square root of the vector dimension of query matrix q.

The updating equation for the knowledge reasoning process is as follows:

$$h_i^{(l+1)} = GELU(f_n(\sum\nolimits_{j \in N_i \cup \{i\}} \alpha_{ji} m_{ji}) + h_i^l) \tag{25}$$

where $h_i^l \in \mathbb{R}^D$ represents the hidden representation of node $v_i \in \mathcal{V}'$ in the layer l model, N_i represents the set of neighboring nodes of node v_i, and $f_n : \mathbb{R}^D \to \mathbb{R}^D$ stands for batch normalization.

3.3.3 Subgraph Reasoning

Subgraph reasoning is the process of knowledge reasoning based on the construction of the disease subgraph and the text subgraph in Section 3.2.

Text subgraph reasoning. Construct the adjacency matrix $\mathcal{A}$ of $\mathcal{T}$:

$$\mathcal{A}_{ij} = \mathcal{A}_{ji} = \begin{cases} 1, (v_i, v_j) \in \mathcal{E}_{sub} \\ 0, (v_i, v_j) \notin \mathcal{E}_{sub} \end{cases}, \forall v_i, v_j \in \mathcal{V}_{sub} \tag{26}$$

A self-loop is added to each node to ensure that each node's representation influences the representation of the subsequent layer. This is accomplished by augmenting the identity matrix I to the base adjacency matrix $\mathcal{A}$ and then normalizing the resulting adjacency matrix $\overline{\mathcal{A}}$ using a degree matrix D:

$$\overline{\mathcal{A}} = \mathcal{A} + I \tag{27}$$

$$D_{ij} = \begin{cases} \sum_k \overline{\mathcal{A}}_{ik}, i = j \\ 0, i \neq j \end{cases} \tag{28}$$

$$\overline{\mathcal{A}} = D^{-\frac{1}{2}} \mathcal{A} D^{-\frac{1}{2}} \tag{29}$$

The reasoning process of $\mathcal{T}$ is as follows:

$$H^{(l+1)} = ReLU(\mathbf{W}_6^{(l)} \overline{\mathcal{A}} H^{(l)} + b_3^{(l)}) \tag{30}$$

where $H^{(l)}$ represents the representation of the $\mathfrak{T}$ in the lth layer. $\mathbf{W}_6^{(l)}$, $b_3^{(l)}$ are the trainable parameters of the lth layer GCN. Reading the $\mathfrak{T}$ in the last layer of the GNN for the whole graph, the output is the vector $\mathcal{K}$, which is the corresponding knowledge reasoning result of $\mathfrak{T}$.

Disease subgraph reasoning. We employ light attention to integrate disease map structure information into the reasoning process. Light attention divides the information into original global features and global features to be updated, where the original global features are defined as follows:

$$\tilde{\mathcal{K}} = mean\ (\mathcal{K} + De) \tag{31}$$

where $\mathcal{K}$ represents the raw knowledge reasoning result, De represents the vector representation of the $\mathfrak{D}$, and $\tilde{\mathcal{K}}$ represents the raw global features.

The global features to be updated are then obtained through the weight matrix $\mathbf{W}$ with $\tilde{\mathcal{K}}$ and De:

$$Com = Concat\ (\tilde{\mathcal{K}} \odot \mathcal{K};\ \tilde{\mathcal{K}} \odot De) \tag{32}$$

$$weight = softmax\ (\mathbf{W}_7 Com + b_4) \tag{33}$$

$$\overline{\mathcal{K}} = weight * Concat\ (\mathcal{K};\ De) \tag{34}$$

where $\mathbf{W}_7$ represents the trainable parameter matrix and b_4 represents the offset. *weight* is a weight representation that measures the importance of $\mathcal{K}$ and De for subsequent updates and is obtained by normalizing the dot product $\tilde{\mathcal{K}}$ with $\mathcal{K}$ and De.

The final output of the knowledge reasoning is obtained by weighted addition of $\overline{\mathcal{K}}$ and $\tilde{\mathcal{K}}$ and employing multi-head light attention computation:

$$\mathcal{M} = \omega_1 \overline{\mathcal{K}} + \omega_2 \tilde{\mathcal{K}} \tag{35}$$

$$\mathcal{K}_i^{'} = \mathbf{W}_8 Concat(\mathcal{M}_0; \mathcal{M}_1; \ldots; \mathcal{M}_{n-1}) + b_5 \tag{36}$$

where $\omega_1 \in [0, 1]$, $\omega_2 \in [0, 1]$ are hyperparameters that measure the relative importance between $\overline{\mathcal{K}}$ and $\tilde{\mathcal{K}}$ in each update. $\mathbf{W}_8$ represents the trainable parameter matrix and b_5 represents the offset.

3.4 Diagnostic Classification

The model can acquire $[T]$, $\mathcal{K}'$ via the text representation (Section 3.1) and the knowledge reasoning (Section 3.3), utilizing interactive attention to amalgamate the output. Specifically, this section can be classified into reasoning result self-attention mechanism (Section 3.4.1), correlation mechanism (Section 3.4.2), and gated fusion mechanism (Section 3.4.3).

3.4.1 Reasoning Result Self-Attention Mechanism

The significance of the knowledge encapsulated within reasoning results varies in influencing the final diagnostic output. To gauge the influence of the embedded knowledge on the final diagnostic result, we introduce the reasoning result self-attention mechanism. The process unfolds as follows:

$$\alpha_i = softmax\ (\mathbf{W}_9 tanh\ (\mathbf{W}_{10} \mathcal{K}_i^{'} + b_6) + b_7) \tag{37}$$

where $\mathbf{W}_9$, $\mathbf{W}_{10}$ represent the trainable parameter matrix and b_6, b_7 represent the offsets. The self-attentive calculation process in reasoning results assigns large weight coefficients to relatively significant knowledge, whereas unimportant knowledge tends towards zero.

3.4.2 Correlation Mechanism

The knowledge noise problem emerges when a sentence veers from its original meaning due to excessive external knowledge introduction. The model gauges the correlation between the text representation and the reasoning result through reasoning result-text representation attention:

$$\beta_i = softmax\ (\mathbf{W}_{11} tanh\ (\mathbf{W}_{12} Concat\ (\mathcal{K}_i^{'}; [T]) + b_8) + b_9) \tag{38}$$

where $\mathbf{W}_{11}$, $\mathbf{W}_{12}$ represent the trainable parameter matrix, and b_8, b_9 represent the offset. β_i represents the weight matrix quantifying the correlation between the text representation and the reasoning result. A stronger correlation between them results in a higher corresponding weight.

3.4.3 Gated Fusion Mechanism

Employing gating mechanisms to integrate text representations and knowledge reasoning results in enhanced mitigation of knowledge noise problems:

$$weight_2 = sigmoid\ (\mathbf{W}_{13} Concat\ (\alpha_i;\ \beta_i) + b_{10}) \tag{39}$$

$$weight_2^{'} = (1 - weight_2) \odot \alpha_i + weight_2 \odot \beta_i \tag{40}$$

$$[T]_{fin} = weight_2^{'} \odot [T] \tag{41}$$

where $\mathbf{W}_{13}$ represents the trainable parameter matrix, b_{10} represents the offset, and $[T]_{fin}$ represents the final hidden representation. $weight_2$ can further measure the degree of influence of the knowledge reasoning results and the text representation on the final output results.

The final prediction is derived by mapping the final hidden representation into the disease space:

$$output = Sigmoid\ (\mathbf{W}_{out} [T]_{fin}) \tag{42}$$

where $\mathbf{W}_{out}$ represents the matrix of trainable parameters.

4 Experiments

4.1 Datasets

Experimental evaluations utilized the multi-labeled Chinese Obstetrics EMRs (**COEMRs**) dataset and the publicly available single-labeled Chinese EMRs (**C-EMRs**) dataset. The Chinese Obstetrics Knowledge Graph (**COKG**) and Chinese Medical Knowledge Graph (**CMeKG**) serve as external knowledge sources for COEMRs and C-EMRs, respectively. Their basic statistics are provided in Tables 1 and 2. Furthermore, detailed information regarding these datasets is presented below:

Table 1: EMR dataset statistics

Dataset	Num	Train	Test	Diseases	Category
COEMRs	24,339	21,905	2434	73	Multiple
C-EMRs	18,331	16,498	1833	10	Single

Table 2: KG dataset statistics

Dataset	Entities	Relations	Used for
COKG	42,616	66,261	COEMRs
CMeKG	68,722	830,768	C-EMRs

- **COEMRs:** A multiple-label dataset comprises real EMRs from multiple hospitals, totaling 24,339 entries. Of these, 21,905 are allocated to the training set and 2434 to the testing set, following a 9:1 division ratio. The dataset retains 73 diseases for diagnosis, constituting a multi-labeled dataset. Specific samples are shown in Fig. 4.

Section	English content	Chinese content
Sex	female	女
Age	thirty-six years old	30岁
Chief complaint	more than 6 months after menopause and 4 hours of vaginal bleeding.	停经6月余，下腹部疼痛4小时
Medical history	This pregnant woman is regular in menstruation, stop menstruating more than 30 days from test urine HCG positive. More than 1 month after menopause, B ultrasound diagnosis of intrauterine early pregnancy. Menopause 40 days, nausea, vomiting and other early pregnancy reactions... ...	该孕妇平素月经规律.停经30天自测尿HCG阳性。停经1月余行B超检查诊断为宫内早孕。停经40天出现恶心、呕吐等早孕反应。
Admitting physical examination	T: 36.6 ° C, P: 80/Min, R: 20/Min, BP: 120/80mmHg, normal development, medium nutrition, conscious, mental can, step into the ward, independent posture, physical examination cooperation. The whole body skin mucous membrane ruddy has no stained yellow, the rash, the bleeding spot, has not touched the swelling superficial lymph node... ...	T:36.6° C, P:80次/分, R:20次/分, BP:120/80mmHg, 发育正常, 营养中等, 神志清, 精神可, 可步入病房, 自主体位, 查体合作。全身皮肤粘膜红润, 无黄染, 皮疹, 出血点, 未触及肿大的浅表淋巴结... ...
Obstetric examination	Extrapelvic measurements IS: 24.0 cm IC: 27.0 cm EC: 19.0 cm TO: 9.0 cm. Uterine height 29.0 cm abdominal circumference 93.0 cm fetal heart rate 144 times / minute fetal weight 2600 G, no contractions	骨盆外测量IS:24.0cm IC:27.0cm EC:19.0cm TO:9.0cm 宫高29.0cm 腹围93.0cm 胎心144次/分 胎儿体重2600g, 无宫缩
Auxiliary examinations	Fetal color doppler ultrasound: BPD: 74.0 mm FL: 53.0 mm Afi: 165.0 mm fetal position: breech position S/D2.2, placental Grade I.	胎儿彩超:BPD:74.0mm FL:53.0mm AFI:165.0mm 胎方位: 臀位S/D 2.2 胎盘级别I
Admitting diagnosis	1. threatened premature labor; 2. placenta previa (borderline); 3. Intrauterine pregnancy 28+2 weeks; 4. G3P1; 5. breech presentation; 6. placenta previa (marginal)	1.先兆早产; 2.前置胎盘(边缘性); 3.孕28+2周; 4.孕3产1; 5.臀位; 6.边缘性前置胎盘
Diagnostic basis	1. Gestation greater than or equal to 28 weeks and less than 37 weeks 2. Minor vaginal bleeding 3. Cervical dilation with ruptured membranes. 4. Ultrasound demonstrates fetal breech presentation in the lower segment of the uterus.	1.妊娠大于等于28周，小于37周 2.阴道少量出血 3.宫颈扩张，胎膜不完整 4.超声显示胎儿臀部位于子宫下段

Figure 4: COEMRs sample. Chief complaint documents the patient's most significant health problem or symptom at the time of the visit. The medical history includes the patient's past illnesses, surgical history, and other relevant information. Admitting physical examination, obstetric examination, and auxiliary examinations document the patient's clinical findings and serve as the primary sources of numerical information. Admitting diagnosis is the physician's diagnosis of the patient, which serves as the categorical label for the dataset

- **C-EMRs** [22]: A single-label dataset comprises 18,331 EMRs. Following a 9:1 division, 16,498 records were assigned to the training set and 1833 to the test set. Due to substantial missing numerical features in C-EMRs, numerical information extraction was omitted, focusing solely on extracting chief complaint information. In contrast to COEMRs, C-EMRs present diagnostic results in a single-label format, covering ten distinct diseases, such as diabetes and hypertension.
- **COKG** [23]: COKG integrates medical knowledge from diverse sources, serving as a domain-specific KG focusing on common obstetric diseases. Its knowledge base comprises obstetrics textbooks, clinical guidelines, specialized medical dictionaries, and web resources, encompassing 42,616 entities and 66,261 relations.

- **CMeKG** [24]: A comprehensive medical domain KG that consolidates medical knowledge from various sources, including clinical documents, textbooks, encyclopedias, and medical standard nomenclatures. It focuses on diseases, drugs, diagnostic techniques, and treatment methods.

4.2 Experimental Settings

4.2.1 Implementation Details

Python is the programming language utilized for the experiment. Deep learning frameworks include PyTorch, DGL (Deep Graph Library), and Transformers. Details of the experimental environment and parameter settings are shown in Table 3. Specifically including operating system, GPU (Graphics Processing Unit), RAM (Random Access Memory), CUDA (Compute Unified Device Architecture), etc.

Table 3: Experimental environment and parameter settings

Parameter	Value	Parameter	Value
Operating system	Ubuntu 22.04	Maximum text length	512
GPU	RTX 3090 Ti	Learning rate	1e−5
RAM	128 G	Dropout	0.3
Python	3.7.11	Batch size	8
Pytorch	1.7.0	Optimizer	Adam
DGL	0.6.1	GAT layers	4
Transformers	3.0.0	Node category	3
CUDA	11.8	Edge relation types	44

4.2.2 Evaluation Metrics

The selected evaluation metrics are commonly used in text classification tasks. Considering that the datasets include both single-label and multi-label instances, the evaluation metrics consist of precision (P), recall (R), and $F1$ scores for single-label classification tasks, while for multi-label classification tasks, hamming loss (HL) and average precision (AP) are employed. The $F1$ score accounts for both P and R, representing a weighted average of both metrics. Specifically, $F1$ scores are categorized into $F1_{micro}$ and $F1_{macro}$, with $F1_{\text{micro}}$ capturing overall P and R across all labels, and $F1_{\text{macro}}$ serving as an average of $F1$ scores across all labels. HL assesses the score of misclassified instance-label pairs, while AP summarizes the weighted average precision at different thresholds, where each threshold's weight is determined by R of the previous threshold.

$$P = \frac{1}{N}\sum_{i=1}^{N}\frac{|f(x_i) \cap y_i|}{|y_i|} \tag{43}$$

$$R = \frac{1}{N}\sum_{i=1}^{N}\frac{|f(x_i) \cap y_i|}{|f(x_i)|} \tag{44}$$

$$F1_{micro} = \frac{2PR}{P+R} \tag{45}$$

$$F1_{macro} = \frac{1}{N}\sum_{i=1}^{N}\frac{2P_iR_i}{P_i + R_i} \tag{46}$$

$$HL = \frac{1}{N}\sum_{i=1}^{N}|f(x_i) \triangle y_i| \tag{47}$$

$$AP = \sum_{n}(R_n - R_{n-1})P_n \tag{48}$$

where $f(x_i)$ represents the predicted label for the ith sample and y_i represents the true label for the ith sample. P_n and R_n represent P and R at the nth threshold, respectively.

4.3 Results Analysis for Comparative Experiment

To demonstrate the effectiveness of GATiT, we analyzed the experimental results. The overall performance of GATiT was first evaluated using several baseline models including Text Recurrent Neural Network (TextRNN) [25], Text Convolutional Neural Network (TextCNN) [26], Text Recurrent Convolutional Neural Network (TextRCNN) [27], TextRNN with Attention (TextRNN+Att) [28], Transformer [29], Deep Pyramid Convolutional Neural Networks (DPCNN) [30], BERT [11], MacBERT [10], KAIE (Knowledge powered Attention and Information Enhanced) [31], GSKN (Graph-based Structural Knowledge-aware Network) [32], DeGCN (Disease enhanced Graph Convolutional Network) [18]. Subsequently, we analyzed disease-specific $F1$ scores to provide further evidence of GATiT's superiority. Tables 4 and 5 show the overall performance of COEMRs and C-EMRs.

Table 4: Diagnostic predictive results of GATiT and baseline models evaluated on COEMRs (Chinese Obstetrics EMRs) evaluated by $F1_{macro}$, $F1_{micro}$, P, R, AP, and HL

Model	$F1_{micro}$(%)↑	P(%)↑	R(%)↑	$F1_{macro}$(%)↑	AP(%)↑	HL ↓
TextRNN [25]	42.87	74.81	30.04	1.17	40.57	0.02732
TextCNN [26]	59.64	83.81	46.29	11.34	68.63	0.02138
TextRCNN [27]	65.81	78.70	56.55	14.76	71.25	0.02005
TextRNN+Att [28]	58.21	78.36	46.31	7.78	65.08	0.02268
Transformer [29]	76.72	80.65	73.17	–	–	0.01982
DPCNN [30]	68.76	80.10	60.23	19.50	74.13	0.01867
BERT [11]	79.74	80.63	78.87	–	–	0.02362
MacBERT [10]	80.50	**86.98**	74.93	31.82	72.53	0.02230
KAIE [31]	80.83	81.44	**80.22**	47.59	81.99	0.01298
GSKN [32]	81.16	84.20	78.33	41.86	82.18	0.01241
DeGCN [18]	81.81	85.95	78.06	45.92	82.65	0.01184
GATiT	**82.32**	86.69	78.37	**50.62**	**83.01**	**0.01167**

Table 5: Diagnostic predictive results of GATiT and baseline models evaluated on C-EMRs (Chinese EMRs) evaluated by $F1_{macro}$, $F1_{micro}$, P, R, AP, and HL

Model	$F1_{micro}$(%)↑	P(%)↑	R(%)↑	$F1_{macro}$(%)↑
TextRNN [25]	64.63	79.97	54.23	30.38
TextCNN [26]	75.53	87.50	66.45	55.24
TextRCNN [27]	76.32	87.01	67.98	57.70
TextRNN+Att [28]	74.70	85.56	66.28	53.76
DPCNN [30]	77.79	83.90	72.50	61.26
BERT [11]	80.17	81.43	78.95	–
MacBERT [10]	89.37	89.85	87.36	88.33
KAIE [31]	89.70	84.50	**93.56**	87.49
GSKN [32]	89.95	87.85	85.92	86.77
DeGCN [18]	91.39	**90.86**	87.40	89.00
GATiT	**92.68**	90.80	89.64	**90.09**

After preprocessing, numerical information extraction, and chief complaint information extraction, the average text length of COEMRs is 371.83. Due to the more severe lack of numerical information in C-EMRs, only chief complaint information enhancement is retained, resulting in an average text length of 352.22 for C-EMRs. The GATiT comprises 153,394,722 trainable parameters, representing a 51.71% increase compared to GSKN. The mean number of entities in the $\mathfrak{T}$ is 33.73, ranging from 2 to 281. Ultimately, 50 entities are retained.

MacBERT is trained on a large-scale corpus, significantly outperforming traditional deep learning models across all metrics, with P of 86.98%. KAIE enhances its ability to extract pertinent information from the BERT model and perform intelligent diagnosis by incorporating medical knowledge in triples. Moreover, KAIE introduces a comprehensive KG spanning various medical domains that are highly relevant to diagnostic tasks. This broader consideration leads to R of 80.22% and 93.56%. Among the evaluation metrics, HL, which evaluates the misclassification score, AP, which summarizes the accuracy at different thresholds, and $F1_{micro}$, the GATiT demonstrates superior performance. GATiT incorporates context nodes containing text information from EMR and employs GAT for knowledge reasoning. Notably, $F1_{macro}$ achieved 50.62%, surpassing KAIE by 3.03%.

4.4 Results Analysis for Across Disease Groups in COEMRs

To further validate GATiT's performance and showcase its superiority effectively, we categorized the total tags based on their occurrence frequency in COEMRS into five groups: 20–30, 31–80, 81–200, 201–1000, and >1000, as illustrated in Fig. 5.

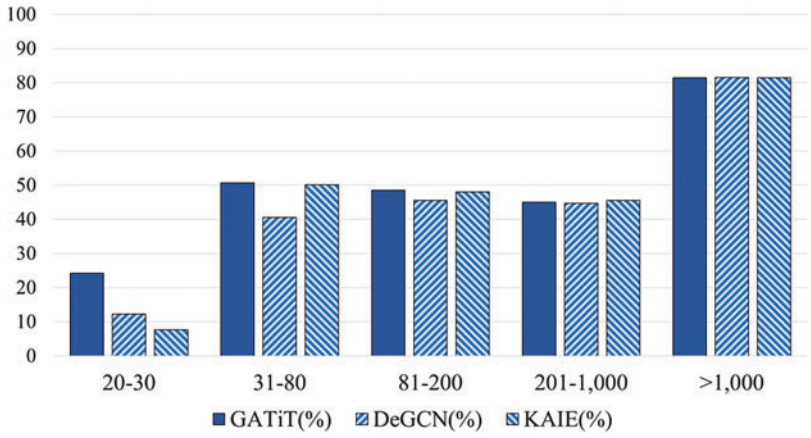

Figure 5: $F1_{micro}$ of GATiT across disease groups in COEMRs

When the label occurrence frequency exceeds 200, the $F1_{micro}$ of GATiT, KAIE, and DeGCN are virtually identical. This observation further validates that with ample data, graph-based knowledge structures and triple representations effectively facilitate the model's learning of disease-related knowledge for diagnosis. GATiT integrates text information from EMRs into the knowledge reasoning process, resulting in more precise reasoning. It outperforms KAIE marginally in the 31–80 and 80–200 frequency groups but slightly lags in the 201–1000 group. For labels with a frequency below 30, the significance of knowledge-based reasoning becomes more pronounced, leading to higher performance for disease-enhanced GATiT compared to KAIE. Moreover, GATiT augmented with EMR representations exhibit further improvement. From a modeling perspective, incorporating context nodes with EMR text data in the reasoning process mitigates errors, thereby reducing noise associated with erroneous reasoning results.

4.5 Results Analysis for Each Disease in C-EMRs

Fig. 6 visualizes the $F1_{micro}$ for each disease in C-EMRs. GATiT demonstrated improvements across five disease categories: "Hypertension", "Diabetes Mellitus", "Gout", "Arrhythmia", and "Urinary Tract Infection (UTI)", compared to DeGCN. Specifically, GATiT exhibited enhancements for "Arrhythmia", "UTI", and the abovementioned categories. Additionally, improvements were observed for "Gastritis", "Gastric Polyps", and "Gastric Ulcer" diseases. Notably, while "Gastric Polyps" showed substantial enhancement in $F1_{micro}$, there was a slight decline in performance for "Gastritis" and "Gastric Ulcer". Nonetheless, the overall average $F1_{micro}$ for these three categories remained higher than that of DeGCN. Moreover, in "Asthma" and "Chronic Obstructive Pulmonary Disease (COPD)", although the $F1_{micro}$ of GATiT were slightly lower than those of DeGCN, they still surpassed those of MacBERT.

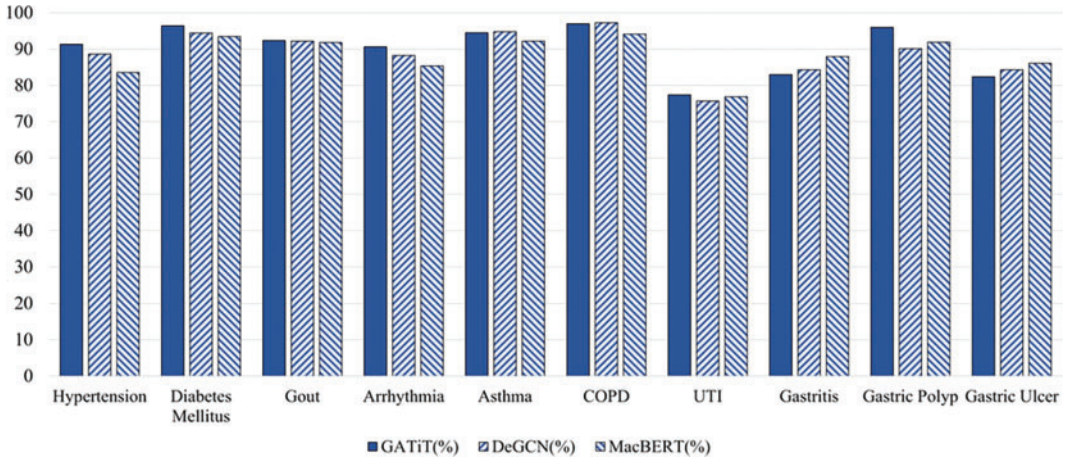

Figure 6: $F1_{micro}$ of GATiT for each disease in C-EMRs

4.6 Results Analysis for the Confusion Matrix

To demonstrate GATiT's predictions across diseases, confusion matrices of GATiT and DeGCN were visualized in Figs. 7 and 8 (rounded to one decimal place, and results of 0 are not shown).

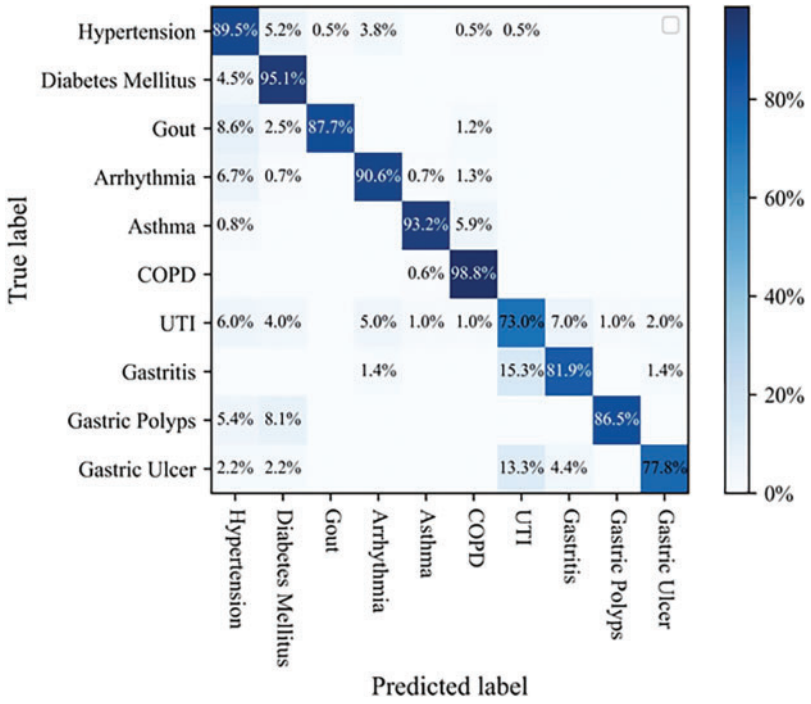

Figure 7: Visualization of DeGCN's confusion matrix in C-EMRs

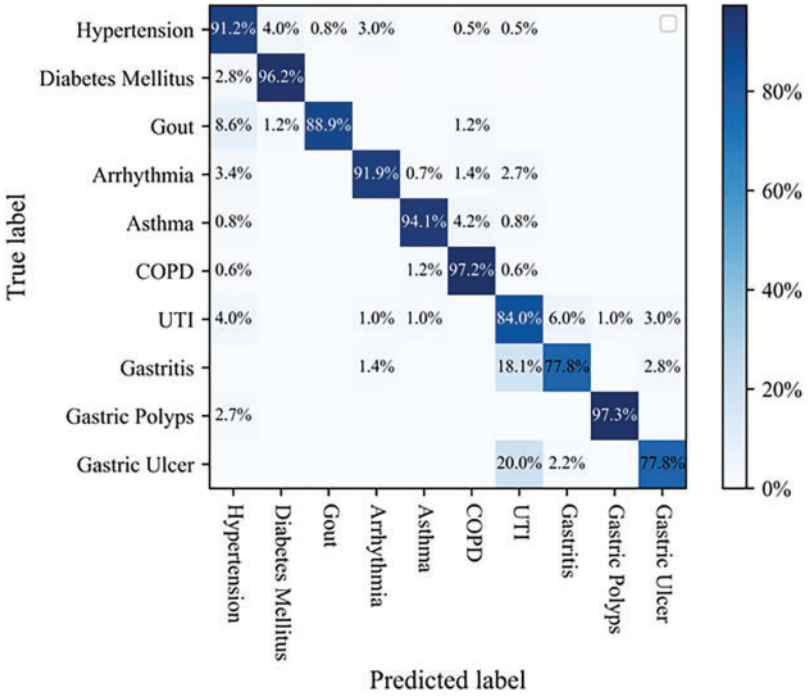

Figure 8: Visualization of GATiT's confusion matrix in C-EMRs

p value of GATiT slightly decreases in "COPD" and "Gastritis", consistent with its $F1_{micro}$ performance. In other diseases, GATiT's p value improves compared to DeGCN. However, for "Asthma" and "Gastric Ulcer", GATiT's $F1_{micro}$ are lower due to increased incorrect predictions of other diseases. Experimental results on C-EMRs further demonstrate the validity of GATiT and the feasibility and necessity of incorporating text information from EMRs into the knowledge reasoning process.

4.7 Results Analysis for Each Disease in COEMRs

To validate the effectiveness of incorporating EMR text nodes and GAT into the knowledge reasoning process, we selected GATiT and DeGCN for further comparison. The specific results were analyzed by examining the $F1_{micro}$ of diseases in 12 specific COEMRs, as shown in Fig. 9. From both the dataset's perspective and the model's perspective, COEMRs exhibit a long-tailed distribution phenomenon.

Integrating the disease subgraph and the text subgraph helps enrich the data on low-frequency diseases, thereby improving knowledge reasoning. Additionally, incorporating EMR text nodes into the knowledge reasoning process allows the model to learn context information from EMRs, thus enhancing diagnostic accuracy.

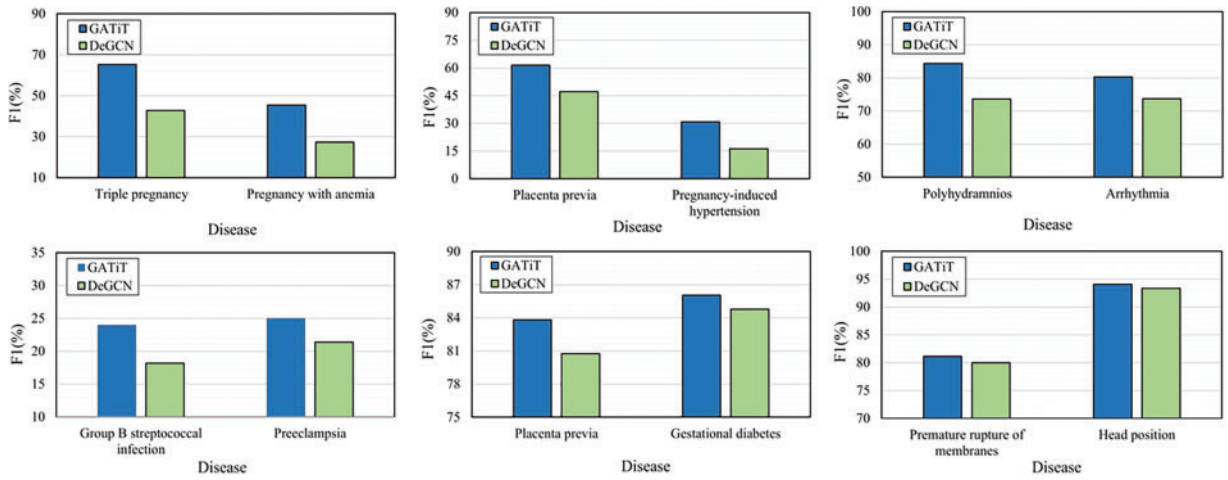

Figure 9: Disease-specific $F1_{micro}$ comparisons in COEMRs

4.8 Results Analysis for Ablation Study

To evaluate the effectiveness of GNN knowledge reasoning, light attention, and the introduced context nodes and additional features, we conducted an ablation study on COEMRs and C-EMRs. The experiments included the following conditions: (a) MHSA (Multi Head Self Attention): utilization of the multi-head self-attention mechanism in the transformer instead of light attention. (b) w/o GCN: removal of the GCN-based knowledge reasoning, with direct reading of the disease subgraph as the reasoning results. (c) w/o Att: removal of light attention. (d) w/o De: elimination of the entire disease augmentation module. (e) w/o NodeZ: removal of the context node containing the EMR information. (f) w/o Ext: elimination of node additional features and edge additional features. Complete removal of additional features leads to the degradation of GATiT to the GAT; thus, w/o Ext can be used to represent knowledge reasoning using an ordinary GAT. Specific experimental results are presented in Tables 6 and 7.

Table 6: The ablation study was conducted in the intelligent diagnostic using COEMRs, and the evaluation metrics included $F1_{macro}$, $F1_{micro}$, P, R, AP, and HL

Model	$F1_{micro}$(%)↑	P(%)↑	R(%)↑	$F1_{macro}$(%)↑	AP(%)↑	HL ↓
GATiT	**82.32**	86.69	**78.37**	**50.62**	**83.01**	**0.01167**
MHSA	81.40	85.64	77.55	44.92	82.96	0.01209
w/o GCN	81.38	85.22	77.88	44.76	82.24	0.01216
w/o Att	81.29	**87.44**	75.95	38.95	81.31	0.01193
w/o De	81.16	84.20	78.33	41.86	82.18	0.01241
w/o NodeZ	81.87	87.20	77.15	45.49	82.63	0.01182
w/o Ext	82.19	86.57	78.23	48.46	82.96	0.01174

Table 7: The ablation study was conducted in the intelligent diagnostic using C-EMRs, and the evaluation metrics included $F1_{macro}$, $F1_{micro}$, P, R, AP, and HL

Model	$F1_{micro}$(%)↑	P(%)↑	R(%)↑	$F1_{macro}$(%)↑
GATiT	**92.68**	**90.80**	**89.64**	**90.22**
MHSA	90.96	89.40	87.68	88.37
w/o GCN	90.93	90.58	87.10	88.60
w/o Att	90.54	88.00	87.93	87.85
w/o De	89.95	87.85	85.92	86.77
w/o NodeZ	91.43	90.57	87.27	89.13
w/o Ext	92.35	90.73	87.53	89.98

The ablation study revealed a significant decrease in all evaluation metrics of GATiT when the knowledge reasoning based on GCNs was removed. This finding underscores the necessity of incorporating knowledge reasoning based on GNNs in the intelligent diagnosis task. Substituting the light attention with the MHSA in the transformer resulted in a decrease in all evaluation metrics of the model, demonstrating the effectiveness of the light attention. The light attention utilized 4,202,512 trainable parameters and required a memory size of 16.81 MB. In contrast, during the experiment, the MHSA employed 16,789,504 trainable parameters and required a memory size of 67.58 MB. Complete removal of the attention mechanism resulted in the model's inability to differentiate between the disease knowledge to be augmented effectively and that embedded in the textual entities, leading to respective decreases of 1.03% and 2.14% in $F1_{micro}$.

In multi-labeled COEMRs, where numerous diseases may be diagnosed in each medical record, removing the attention mechanism caused an 11.67% decrease in the $F1_{macro}$ evaluation metric for GATiT. In the single-label C-EMRs dataset, with a small number of independent labels, combining the disease subgraph without relying on the attentional mechanism can effectively capture knowledge related to the diagnosed disease. Removing the context nodes of the EMRs led to a decrease in the model's performance relative to GATiT, which is attributed to the sparser nature of the COKG, where the added node and edge features only moderately influence the representation of the original nodes. The ablation mentioned above experiments further emphasizes the need to introduce and distinguish context nodes from other nodes in the knowledge reasoning process.

5 Conclusion

This paper proposes an innovative GATiT model, which incorporates the graph attention network in knowledge reasoning for medical intelligent diagnosis tasks. GATiT incorporates the disease subgraph related to the diagnosis and constructs the text subgraph of context nodes containing text information from EMRs within the knowledge reasoning. To facilitate the model's ability to differentiate context nodes from other nodes during learning, GATiT introduces additional features. Additionally, we adapt the message-passing strategy of GAT and adjusts the calculation of attention weights to learn these features. In the text representation, we adjust MacBERT to obtain text representations of the EMR. In the diagnostic classification, we use the interactive attention-based fusion mechanism to fuse the above results to get the final output. Results from COEMRs and C-EMRs validate the effectiveness of GATiT and the importance of incorporating text information from EMRs in the knowledge reasoning process. For future work, we plan to:

- Further explore how to effectively utilize different forms of multimodal data in intelligent diagnosis research and realize the effective fusion of different modal data.
- Another focus of future work is to apply the few-shot learning technique to intelligent diagnosis research to learn the relevant features of diseases from a small amount of data.

Acknowledgement: We thank all the anonymous reviewers who generously contributed their time and efforts. Their professional recommendations have greatly enhanced the quality of the manuscript.

Funding Statement: This work was supported in part by the Science and Technology Innovation 2030–"New Generation of Artificial Intelligence" Major Project (No. 2021ZD0111000), and Henan Provincial Science and Technology Research Project (No. 232102211039).

Author Contributions: The authors confirm contribution to the paper as follows: study conception and design: Yu Song, Pengcheng Wu; data collection: Kunli Zhang, Yu Song; analysis and interpretation of results: Pengcheng Wu, Dongming Dai; draft manuscript preparation: Kunli Zhang, Pengcheng Wu, Mingyu Gui. All authors reviewed the results and approved the final version of the manuscript.

Availability of Data and Materials: A dataset used in this paper can be found in website. CMeKG can be accessed on http://nscc.zzu.edu.cn/know/ (accessed on 26 March 2024). The datasets C-EMRs can be downloaded from the website https://github.com/YangzlTHU/C-EMRs (accessed on 20 April 2018). Other data is available on request from the authors.

Ethics Approval: Not applicable.

Conflicts of Interest: The authors declare that they have no conflicts of interest to report regarding the present study.

References

[1] J. H. Kamdar, J. Jeba Praba, and J. J. Georrge, "Artificial intelligence in medical diagnosis: Methods, algorithms and applications," *Mach. Learn. Health Care Perspect. Mach. Learn. Healthc.*, vol. 13, pp. 27–37, Mar. 2020. doi: 10.1007/978-3-030-40850-3_2.

[2] C. C. Aggarwal and C. Zhai, "A survey of text classification algorithms," in *Mining Text Data*, 6th ed. Boston, MA, USA: Springer, 2012, pp. 163–222. doi: 10.1007/978-1-4614-3223-4_6.

[3] K. Zhang, H. Ma, Y. Zhao, H. Zan, and L. Zhuang, "The comparative experimental study of multilabel classification for diagnosis assistant based on Chinese obstetric EMRs," *J. Healthc. Eng.*, vol. 2018, no. 1, Feb. 2018, Art. no. 7273451. doi: 10.1155/2018/7273451.

[4] S. Minaee, N. Kalchbrenner, E. Cambria, N. Nikzad, M. Chenaghlu and J. Gao, "Deep learning-based text classification: A comprehensive review," *Assoc. Comput. Mach.*, vol. 54, no. 62, pp. 1–40, Apr. 2021. doi: 10.1145/3439726.

[5] A. Hogan *et al.*, "Knowledge graphs," *ACM Comput. Surv.*, vol. 54, no. 71, pp. 1–37, Jul. 2021. doi: 10.1145/3447772.

[6] X. Hao *et al.*, "Construction and application of a knowledge graph," *Remote Sens.*, vol. 13, no. 13, Jun. 2021, Art. no. 2511. doi: 10.3390/rs13132511.

[7] X. Chen, S. Jia, and Y. Xiang, "A review: Knowledge reasoning over knowledge graph," *Expert. Syst. Appl.*, vol. 141, no. 1, Mar. 2020, Art. no. 112948. doi: 10.1016/j.eswa.2019.112948.

[8] F. Scarselli, M. Gori, A. C. Tsoi, M. Hagenbuchner, and G. Monfardini, "The graph neural network model," *IEEE Trans. Neural Netw.*, vol. 20, no. 1, pp. 61–80, Dec. 2008. doi: 10.1109/TNN.2008.2005605.

[9] S. Sun *et al.*, "Review of graph neural networks applied to knowledge graph reasoning," *J. Front. Comput. Sci. Technol.*, vol. 17, no. 1, pp. 27–52, Jul. 2023.
[10] Y. Cui, W. Che, T. Liu, B. Qin, S. Wang and G. Hu, "Revisiting pre-trained models for Chinese natural language processing," in *Findings of the Assoc. for Comput. Linguist.: EMNLP 2020*, Nov. 2020, pp. 657–668. doi: 10.18653/y1/2020.findings-emnlp.58.
[11] J. Devlin, M. Chang, K. Lee, and K. Toutanova, "BERT: Pre-training of deep bidirectional transformers for language understanding," in *Proc. of the 2019 Conf. of the North Am. Chapter of the Assoc. for Comput. Linguist.: Hum. Lang. Technol.*, Minneapolis, Minnesota, Jun. 2019, pp. 4171–4186. doi: 10.18653/v1/n19-1423.
[12] P. Veličković, G. Cucurull, A. Casanova, A. Romero, P. Liò and Y. Bengio, "Graph attention networks," in *6th Int. Conf. on Learn. Representat.*, Vancouver, BC, Canada, Apr. 2018.
[13] X. Li, H. Wang, H. He, J. Du, J. Chen and J. Wu, "Intelligent diagnosis with Chinese electronic medical records based on convolutional neural networks," *BMC Bioinformatics*, vol. 20, no. 1, Feb. 2019, Art. no. 62. doi: 10.1186/s12859-019-2617-8.
[14] Y. Sun, J. Liu, K. Yu, M. Alazab, and K. Lin, "PMRSS: Privacy-preserving medical record searching scheme for intelligent diagnosis in IoT healthcare," *IEEE Trans. Ind. Inform.*, vol. 18, no. 3, pp. 1981–1990, Mar. 2022. doi: 10.1109/TII.2021.3070544.
[15] J. Chen, Q. Yuan, C. Lu, and H. Huang, "A novel sequence-to-subgraph framework for diagnosis classification," in *Proc. of the Thirtieth Int. Joint Conf. on Artif. Intell.*, Montreal, QC, Canada, Aug. 2021, pp. 3606–3612. doi: 10.24963/ijcai.2021/496.
[16] K. Yang *et al.*, "KerPrint: Local-global knowledge graph enhanced diagnosis prediction for retrospective and prospective interpretations," *Proc. AAAI Conf. Artif. Intell.*, vol. 37, no. 4, pp. 5357–5365, Jun. 2023. doi: 10.1609/aaai.v37i4.25667.
[17] X. Yang, Y. Zhang, F. Hu, Z. Deng, and X. Zhang, "Feature aggregation-based multi-relational knowledge reasoning for COPD intelligent diagnosis," *Comput. Electr. Eng.*, vol. 114, 2024, Art. no. 109068. doi: 10.1016/j.compeleceng.2023.109068.
[18] Y. Song *et al.*, "Research on double-graphs knowledge-enhanced intelligent diagnosis," in *China Health Inf. Process. Conf.*, Hangzhou, China, Sep. 2023, pp. 317–332. doi: 10.1007/978-981-99-9864-7_21.
[19] C. Qu, L. Yang, M. Qiu, W. B. Croft, Y. Zhang and M. Iyyer, "BERT with history answer embedding for conversational question answering," in *Proc. of the 42nd Int. ACM SIGIR Conf. on Res. and Dev. in Inf. Retrieval*, Paris, France, Jul. 2019, pp. 1133–1136. doi: 10.1145/3331184.3331341.
[20] Y. Lin, Z. Liu, M. Sun, Y. Liu, and X. Zhu, "Learning entity and relation embeddings for knowledge graph completion," in *Proc. AAAI Conf. Artif. Intell.*, vol. 29, no. 1, Feb. 2015. doi: 10.1609/aaai.v29i1.9491.
[21] Y. Liu *et al.*, "RoBERTa: A robustly optimized BERT pretraining approach," 2019, *arXiv: 1907.11692*.
[22] Z. Yang, Y. Huang, Y. Jiang, Y. Sun, Y. Zhang and P. Luo, "Clinical assistant diagnosis for electronic medical record based on convolutional neural network," *Sci. Rep.*, vol. 8, no. 1, p. 6329, Apr. 2018. doi: 10.1038/s41598-018-24389-w.
[23] K. Zhang, C. Hu, Y. Song, H. Zan, Y. Zhao and W. Chu, "Construction of Chinese obstetrics knowledge graph based on the multiple sources data," in *Workshop on Chinese Lexical Semantics*, Nanjing, China, May 2021, pp. 399–410. doi: 10.1007/978-3-031-06547-7_31.
[24] Aodema *et al.*, "Preliminary exploration of constructing Chinese medical knowledge graph CMeKG," *J. Chin. Inf. Process.*, vol. 33, no. 10, pp. 1–7, 2019.
[25] S. Hochreiter and J. Schmidhuber, "Long short-term memory," *Neural Comput.*, vol. 9, no. 8, pp. 1735–1780, Nov. 1997. doi: 10.1162/neco.1997.9.8.1735.
[26] Y. Kim, "Convolutional neural networks for sentence classification," in *Proc. of the 2014 Conf. on Empirical Methods in Nat. Lang. Process. (EMNLP)*, Doha, Qata, Oct. 2014, pp. 1746–1751. doi: 10.3115/v1/D14-1181.
[27] S. Lai, L. Xu, K. Liu, and J. Zhao, "Recurrent convolutional neural networks for text classification," in *Proc. AAAI Conf. Artif. Intell.*, vol. 29, no. 1, pp. 2267–2273, 25–30 Jan., 2015.
[28] P. Zhou *et al.*, "Attention-based bidirectional long short-term memory networks for relation classification," in *Proc. of the 54th Annu. Meeting of the Assoc. for Comput. Linguist.*, Berlin, Germany, 2016, pp. 207–212.

[29] A. Vaswani *et al.*, "Attention is all you need," in *Adv. in Neural Inf. Process. Syst.*, Long Beach, CA, USA, 4–9 Dec. 2017, pp. 6000–6010.
[30] R. Johnson and T. Zhang, "Deep pyramid convolutional neural networks for text categorization," in *Proc. of the 55th Annu. Meeting of the Assoc. for Comput. Linguist.*, Vancouver, BC, Canada, Jul. 2017, pp. 562–570. doi: 10.18653/v1/P17-1052.
[31] K. Zhang, X. Zhao, L. Zhuang, Q. Xie, and H. Zan, "Knowledge-enabled diagnosis assistant based on obstetric EMRs and knowledge graph," in *China Nat. Conf. on Chin. Comput. Linguist.*, Haikou, China, Springer, Oct. 2020, pp. 444–457. doi: 10.1007/978-3-030-63031-7_32.
[32] K. Zhang, B. Hu, F. Zhou, Y. Song, X. Zhao and X. Huang, "Graph-based structural knowledge-aware network for diagnosis assistant," *Math. Biosci. Eng.*, vol. 19, no. 10, pp. 10533–10549, Jul. 2022. doi: 10.3934/mbe.2022492.