IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received August 5, 2020, accepted August 11, 2020, date of publication August 14, 2020, date of current version August 25, 2020.

Digital Object Identifier 10.1109/ACCESS.2020.3016676

Diagnosis Method of Thyroid Disease Combining Knowledge Graph and Deep Learning

XUQING CHAI
College of Computer and Information Engineering, Henan Normal University, Xinxiang 453000, China
e-mail: chaixuqing@htu.edu.cn

This work was supported in part by the National Natural Science Foundation of China under Grant U1804164, in part by the Key Project of Science and Technology of Henan Provincial Science and Technology Department under Grant 192102310020, in part by the Henan Provincial Federation of Social Sciences under Grant SKL-2018-771, in part by the Research Project on Curriculum Reform of Teacher Education in Henan Province under Grant 2018-JSJYYB-020, in part by the Education Science Research Fund of Henan Normal University under Grant 2018JK10, in part by the High Performance Computing Centre of Henan Normal University, and in part by the Supercomputing Center of University of Science and Technology of China.

ABSTRACT The scale of medical data is growing rapidly, and these data come from different data sources. The amount of data is huge, the production speed is fast, and the format is different. Case data is very important because it contains a lot of medical knowledge about diseases, drugs, treatments, etc. It can provide important support for the development of smart medicine. Knowledge graph is a graph-based data structure, which can well represent the relationship between these medical data in reality and form a semantic network. This research uses knowledge graph technology to connect trivial and scattered knowledge in various medical information systems to assist in disease diagnosis. This research takes thyroid disease as an example, constructs a medical knowledge graph and applies it to intelligent medical diagnosis. First, extract the relationships between biomedical entities to construct a biomedical knowledge graph. Then, the entities and relationships in the knowledge graph are transformed into low-dimensional continuous vectors through the knowledge graph embedding method. Finally, the known pathological disease relationship data is used to train the disease diagnosis model of the bidirectional long short-term memory network (BSTLM). Experiments show that the thyroid disease diagnosis method that combines knowledge graphs and deep learning has a better diagnostic effect. This shows that smart medical care based on the knowledge graph will provide a solution path for alleviating the shortage of domestic high-quality medical resources.

INDEX TERMS Knowledge graph, deep learning, thyroid disease, disease diagnosis.

I. INTRODUCTION

Intelligent medical care [1], [2] refers to the use of technologies such as the Internet of Things and artificial intelligence to realize the interaction between patients and medical staff, medical institutions, and medical equipment, and gradually achieve informatization. Use information technology to improve disease prevention, diagnosis and research, so as to achieve scientific management of population health. Ultimately benefit all components of the medical ecosystem. The core and key of intelligent medical care is intelligent diagnosis and treatment, that is, making computers become a brain with medical knowledge, so as to provide assistant decision-making for doctors' diagnosis and treatment. Such computers are capable of diagnosis and treatment. It can not only independently provide medication assistance, triage guidance, health consultation and other services. It can also assist medical practitioners in completing certain tasks with high quality.

Personal information such as age and gender of the patient is recorded in the EMR electronic medical record. The EMR electronic medical record records the patient's age, gender and other personal information, as well as treatment information such as the diagnosis results of each treatment, the length of stay in the hospital, and the medication status. Each classification information can be regarded as an entity, and these entities are related to each other and have different relationships. Using the knowledge graph to describe the relevant medical knowledge in the EMR electronic medical record can improve its utilization rate, promote the development of

The associate editor coordinating the review of this manuscript and approving it for publication was Yizhang Jiang.

intelligent medical care, and play an important auxiliary role in providing decision support for doctors [3]. Since the knowledge graph is superior to the expressive ability of general relational databases, it plays an increasingly important role in the process of processing these massive medical data. The knowledge graph can integrate isolated data. As a structured graph model, it can well describe the relationship between various entities in reality.

The knowledge graph [4], [5] is a relational network that connects different information together. At present, the more mature large-scale general knowledge graphs mainly include YAGO [6], DBpedia [7], NELL [8], Freebase [9]. Knowledge graphs have been widely used in education [10], [11], medical treatment [12], [13] and other fields. Knowledge graph is an important technology in the field of artificial intelligence, and it is the technical foundation for building a computer medical knowledge brain. Therefore, the knowledge graph has become one of the key technologies of smart medicine. After Ledley [14] and others first applied the data model to the field of clinical medicine, various forms of medical expert assistance systems appeared. The main workflow of these systems is to structure the clinical experience and knowledge of medical experts to establish a medical knowledge base. Then make inference rules through experts. Finally, in practical applications, diagnosis and reasoning are performed based on the medical examination data input by the user. However, the mechanization of this system and the overly simple rule-based reasoning method have certain limitations in constructing knowledge bases and diagnostic reasoning for medical data with diverse data. With the development of computer technology, machine learning, and artificial intelligence technology [15]–[23], more and more scholars have begun to use machine learning and artificial intelligence technology to build knowledge bases and disease-aided diagnosis systems. Since the 1970s, foreign countries have invested a lot of manpower and material resources in this area of research and development, and have achieved fruitful results [24]–[31].

Most of the above applications based on knowledge graphs in the medical field use traditional machine learning [32]–[40]. The recognition rate of this method needs to be further improved. Therefore, this research introduces a deep learning algorithm called BLSTM [41] to be used in the diagnosis of thyroid diseases. First, construct a knowledge map of thyroid diseases. Second, extract the entities and relationships related to the thyroid in the knowledge graph, and use the knowledge graph embedding to convert them into low-dimensional continuous vectors. Finally, train the BLSTM diagnostic model. Input the characteristic word vector of the thyroid gland and the relevant knowledge entity vector into the trained model to obtain the decision result. The main work of this paper is summarized as follows

(1) Constructed a knowledge map for the thyroid. Extract useful information from the thyroid patient information database, thyroid examination information database, thyroid drug use information database, and thyroid index information database. According to the connection between entities, the BFS breadth-first algorithm [42] is used to fill the created concept tree to obtain the thyroid knowledge graph.

(2) Train the thyroid diagnostic model BLSTM by using the constructed knowledge map. Through experimental comparison, thyroid disease diagnosis based on deep learning and knowledge map fusion is more efficient than manual diagnosis and more accurate than diagnosis based on machine learning methods. The classification performance based on BLSTM shows better results than other deep learning algorithms.

II. RELATED WORK

A. KNOWLEDGE GRAPH

The knowledge graph is a semantic network that describes the entities, concepts, and their relationships in the real world. It fully uses visualization technology, not only can describe knowledge resources and carriers, but also can analyze and describe knowledge and the connections between knowledge. In the past, most researches on medical knowledge graphs were based on literature, and most of the knowledge came from public medical literature, and electronic medical record data was rarely used. However, the electronic medical record data covers the whole process of diagnosis and treatment of patients in various departments of the hospital, and has a wealth of medical knowledge. The information in the electronic medical record is more disorderly than the information in the subject area, and most of the data belongs to unstructured text. Perform semantic analysis on the information in the electronic medical record, and extract the knowledge unit used to draw the knowledge graph. Find out the connections between the knowledge units, and use the knowledge graph technology to connect the scattered and fragmented knowledge in the medical resources—electronic medical records. This can provide patients and doctors with comprehensive services. In addition, medical knowledge graphs are widely used in medical information search engines, medical question and answer systems, and medical decision support systems.

The drawing of knowledge graph mainly includes three links: constructing knowledge unit, constructing unit relationship, and structured display of knowledge graph. The extraction of knowledge units and the recognition of relationships between knowledge are mapped to the recognition of named entities and the recognition of entity relationships. The main links of the construction process are shown in **FIGURE 1**.

B. DISEASE DIAGNOSIS BASED ON KNOWLEDGE GRAPH

The main entities included in the medical knowledge map are diseases, symptoms, parts, departments, and drugs, as well as the relationships between entities. The disease diagnosis and principle diagram based on the knowledge map is shown in **FIGURE 2.**

The user first selects the location based on gender, and then selects their symptoms based on the location. Combining the age, gender and other information selected by the user, based on the reasoning model, seven diseases that the user may have

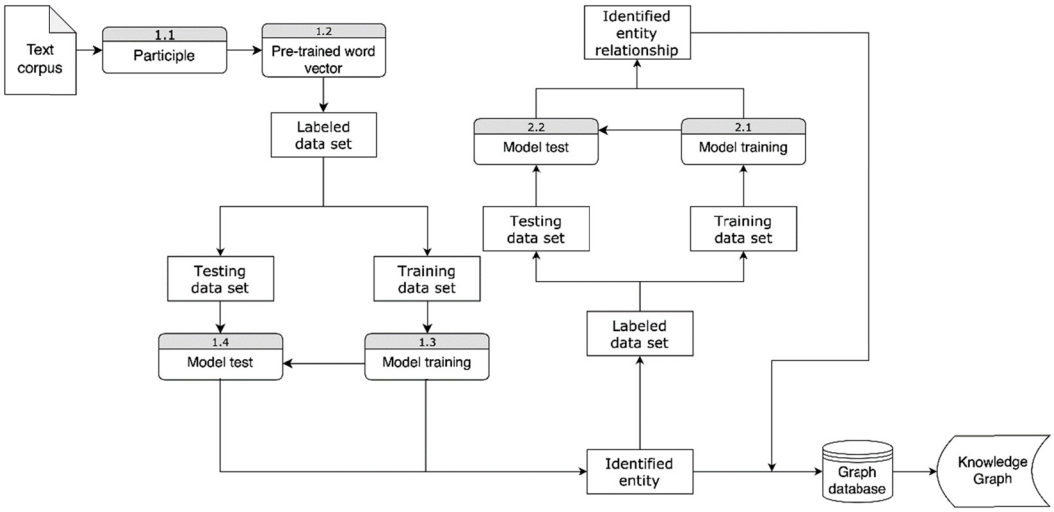

FIGURE 1. Construction process of knowledge graph.

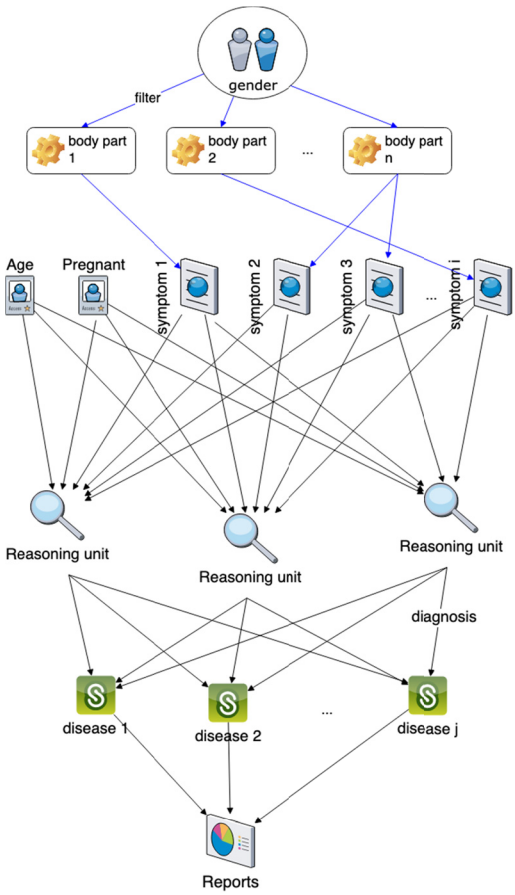

FIGURE 2. Schematic diagram of disease diagnosis based on knowledge graph.

are diagnosed. And show detailed information about each disease, such as overview, etiology, symptoms, complications, and treatment. In the process of disease diagnosis, since the input information is that the user selects the characteristic information according to his own situation, there may be cases of missed selection and wrong selection, especially for symptom characteristics. In the disease diagnosis model, the symptom features have the greatest impact on the diagnosis results, and it is also the place where users may make mistakes in selecting features. Merely describing the relationship between entities in the existing medical knowledge graph is not enough to solve the above problems, so the relationship between disease and symptoms needs to be redefined. Quantify the weights of diseases and symptoms to distinguish different symptoms that have different effects on different diseases. At the same time, based on the knowledge map to find the hidden relationship between diseases and symptoms, in order to reduce the error caused by the uncertainty of the user's selection characteristics.

III. THYROID DISEASE DIAGNOSIS PROCESS

A. THYROID DIAGNOSTIC FRAMEWORK

This article diagnoses whether the input sample has thyroid disease by inputting the medical knowledge graph. The diagnostic model uses BLSTM. The framework of the proposed thyroid disease diagnosis method is shown in **FIGURE 3.**

Firstly, SemRep Based Knowledge Graph (SemKG) is constructed using the entity relationship in Biomedical Abstracts. Then use Knowledge Graph Embedding to convert the entities and relationships in SemKG into low-dimensional continuous vectors. Then use the ''pathology-disease'' data to train Bidirectional Long Short-Term Memory Networks (BLSTM). Finally, use the trained deep learning model combined with the knowledge map to diagnose thyroid diseases.

B. CONSTRUCTION OF THYROID KNOWLEDGE GRAPH

1) BUILD PROCESS

The overall process of this design is: First, according to different data in different tables in the database, combined with the initial thyroid nodule medical records, the conceptual layer of

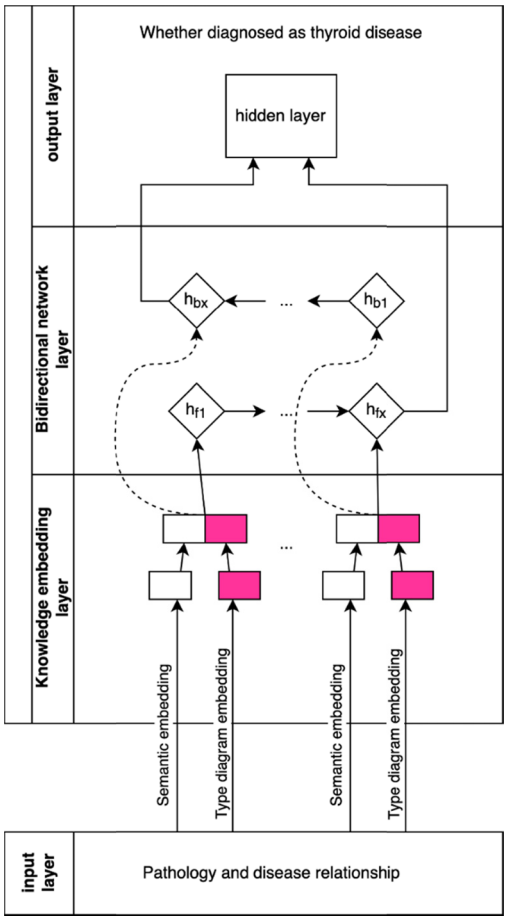

FIGURE 3. Framework diagram of the proposed thyroid disease diagnosis method.

the thyroid knowledge map is extracted. Construct a conceptual classification tree and extract relationships between data. Then, the data in the table, that is, the entities are filled into the conceptual layer. In the form of triples, namely <entity, relationship, entity>, a complete thyroid knowledge map is obtained.

2) CONCEPTUAL DESIGN

The thyroid knowledge map is constructed based on the thyroid disease information stored in the database of a third-class hospital. Including patient information entity, drug use information entity, diagnosis data entity, etc. There are many connections between entities. Since the data in the database is tidy, the relationship between entities and entities can be standardized and analyzed to form the entire knowledge graph. Using the formed knowledge graph to provide semantic relationships, users can directly observe the connection between entity data and entities. The data in the thyroid database is classified, and the following medical entity definition is obtained.

Definition 1: Thyroid medical entity. Including thyroid patient entity, basic information entity, thyroid diagnosis result entity, thyroid medication entity, etc.

TABLE 1. The types of thyroid factual relationships.

relationship	Explanation
X has Y	Entity Y belongs to Entity X
X attribute_of Y	Entity Y has an attribute entity X
X use drug Y	Illness entity X Drug entity Y
X diagnosis Y	The diagnosis result of inspection report entity Y is entity X, and there is a diagnosis relationship between entities.

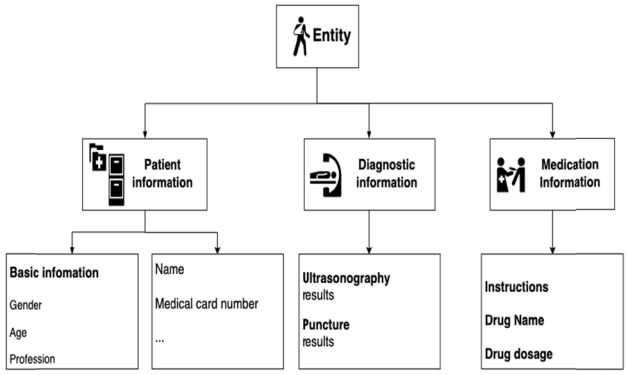

FIGURE 4. The conceptual layer of the thyroid knowledge graph.

Definition 2: Thyroid factual relationship entity. The thyroid factual relationship represents the connection between different thyroid medical entities, such as <patient, hospitalized diagnosis, goiter>. Where the patient and goiter are all thyroid medical entities. Admission diagnosis is thyroid factual relationship entity. The types of thyroid factual relationships are shown in Table 1.

After defining the thyroid medical entity and the thyroid factual relationship entity, the concept classification tree of the thyroid knowledge graph is constructed as shown in **FIGURE 4**.

3) PHYSICAL LAYER FILLING

Through the method of entity mapping, the concepts in the conceptual layer are mapped to the entities in the database. This paper uses the BFS breadth-first algorithm to fill the created concept tree to get the knowledge graph. Input the implemented concept classification tree T, the concept collection C in the concept layer, and the defined entity collection E. Output the thyroid knowledge graph G after operation. It is guaranteed that the output knowledge graph is constructed in the form of triples. The process of entity filling can be expressed as: First, create a mapping table. According to the principle that the entity belongs to a concept in the concept tree, the mapping table between the concept and the entity is constructed, as shown in **FIGURE 5.**

Secondly, fill in entities according to the mapping table. Store the entities in the mapping table in the corresponding sub-nodes through BFS breadth-first traversal, so that each entity has its own attributes and attribute values. Finally,

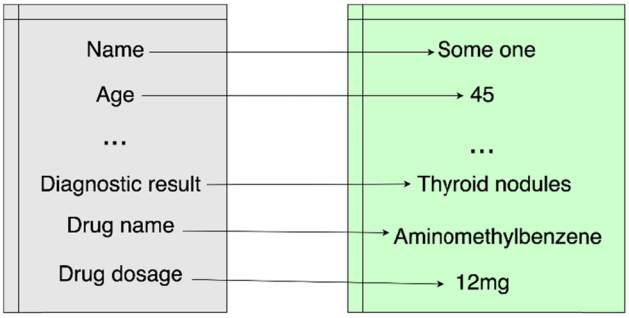

FIGURE 5. Mapping table example.

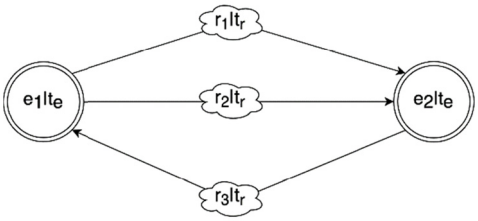

FIGURE 6. Schematic diagram of SemKG.

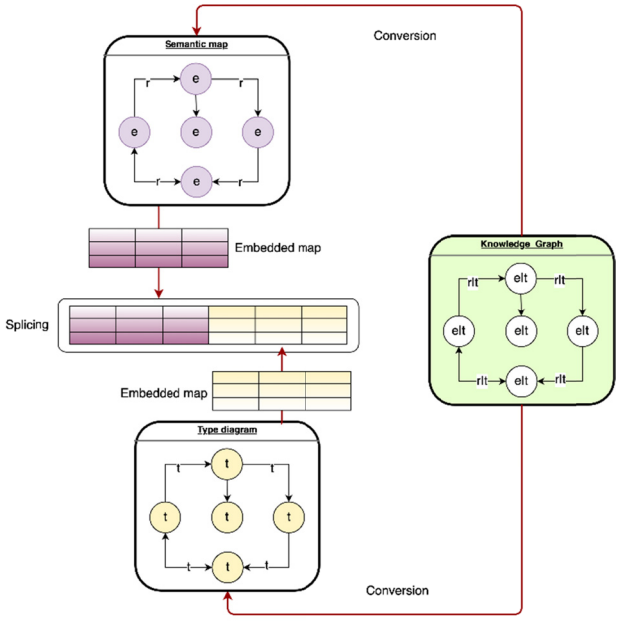

FIGURE 7. Vector transformation process.

extract entities and relationships, comprehensively determine the triples, and form the final thyroid knowledge map. The knowledge graph is expressed as

$$G_{KG} = (E \cup \phi(E), R \cup \phi(R))$$

where $E = \{e_1, e_2, \ldots, e_N\}$ represents N entities in the knowledge graph. $R = \{r_1, r_2, \ldots, r_M\}$ represents the relationship between entities. $T = \{t_1, t_2, \ldots, t_k\}$ represents the semantic type. Each entity e or relationship r can correspond to the semantic type through the relationship mapping function ϕ. **FIGURE 6** is a schematic diagram of SemKG.

C. DIAGNOSIS MODEL OF THYROID DISEASE

Given a path $\pi_i^l = e_0 r_0 e_1 r_{1\ldots} e_{l-1} r_{l-1}$. Where e_0 indicates a certain pathology. e_l indicates whether it is diagnosed as thyroid disease. The goal of the disease diagnosis model is to diagnose the probability that a certain pathology causes thyroid disease. The expression is as follows

$$p\left(y|\pi_i^l\right) = D\left(g\left(\pi_i^l\right), \theta\right)$$

where $D(\cdot)$ represents any discriminant model with parameter θ. $g(\cdot)$ represents the feature extraction function. The input layer of the disease diagnosis model is an arbitrary path $\pi_i^l = e_0 r_0 e_1 r_{1\ldots} e_{l-1} r_{l-1}$. e is the entity, and r is the relationship between the two entities. In the knowledge graph embedding layer, each element x_i in π_i^l is transformed into a vector representation. The conversion process is shown in **FIGURE 7**.

First, transform the knowledge graph $G_{KG} = \left(E \cup \phi(E), R \cup \phi(R)\right)$ into Semantic Graph and Type Graph at the same time. Semantic Graph $G_{SG} = (E, R)$ only contains entities and relationships between entities. Type Graph $G_{TG} = (\phi(E), \phi(R))$ only contains semantic types corresponding to entities and relationships. Translating Embedding (TransE) is used to embed G_{SG} and G_{TG} respectively. Therefore, each element x_i in π_i^l is transformed into a vector. The expression of vector x_i is as follows

$$x_i = g(x_i)_{SG} \prod g(x_i)_{TG}$$

where $g(\cdot)$ represents the knowledge graph embedding method. $\prod$ represents the splicing operation of vectors. The embedded representation of the knowledge graph is obtained by minimizing the loss function. Take the semantic graph as an example.

$$L = \sum_{(e_1,r,e_2)\in S} \sum_{(e_{1'},r,e_{2'})\in S'} [\gamma + d(\mathbf{e}_1 + \mathbf{r}, \mathbf{e}_2) - d(\mathbf{e}_{1'} + \mathbf{r}, \mathbf{e}_{2'})]_+$$

where bold represents the vector representation of the corresponding element. For example, $\mathbf{e}_1$ is the vector of entity e_1. $[x]_+$ represents the part where x is greater than zero. d is the L_1 paradigm. The positive training set $S_{(e_1|r,e_2)}$ contains all triples $(e_1|r, e_2)$ in the semantic graph. The set of negative examples S' is obtained by replacing the entity e_1 or e_2 of the triple in $S_{(e_1|r,e_2)}$. The data embedded in the constructed learning graph is as follows

$$S'_{(e_1,r,e_2)} = \{(e_{1'}, r, e_2) | e_{1'} \in E\} \cup \{(e_1, r, e_{2'}) | e_{2'} \in E\}$$

This research uses Stochastic Gradient Descent (SGD) to get the final graph embedding representation. The graph embedding learning process of type graph is the same as that of semantic graph. The knowledge graph embedding

layer converts each element in π_i^l into a vector of length $L_{SG} + L_{TG}$. Finally, π_i^l is transformed into a matrix $\mathbf{X}$ of size $(L_{SG} + L_{TG}) \times l$.

$$\mathbf{X}_{\pi_i^l} = \bigcup_{x_i \in \pi_i^l} x_i$$

This study uses dual BLSTM to predict the relationship between pathology and disease. The LSTM structure including the input layer x_i, the hidden layer h_i and the output layer y_i is as follows.

$$\begin{aligned} i_t &= \sigma\left(W_{xi}x_t + W_{hi}h_{t-1} + b_i\right) \\ f_t &= \sigma\left(W_{xf}x_t + W_{hf}h_{t-1} + b_f\right) \\ o_t &= \sigma\left(W_{xo}x_t + W_{ho}h_{t-1} + b_o\right) \\ g_t &= \tanh\left(W_{xg}x_t + W_{hg}h_{t-1} + b_g\right) \\ c_t &= f \odot c_{t-1} + i_t \odot g_t \\ h_t &= o \odot \tanh\left(c_t\right) \end{aligned}$$

σ is logistic sigmoid function. i represents the input gate with the same dimension as the hidden layer vector h. f represents the forget gate vector. o represents the output gate vector. c represents the cell activation vector. W_* and b_* are trainable parameters. $\odot$ is a bit multiplication operation. c_0 is the e_0 vector representation in input π_i^l. e_0 represents a certain pathology. The disadvantage of traditional LSTM is that only text information is used, and the order of entities in π_i^l affects the relationship between pathology and disease. Therefore, this study uses BLSTM for thyroid diagnosis. In **FIGURE 3**, BLSTM uses two hidden layers to process data in different directions. The expression is as follows

$$\begin{aligned} h_{ft} &= H\left(W_{xh_f}x_t + W_{h_fh_f}h_{f_{t-1}} + b_{h_f}\right) \\ h_{bt} &= H\left(W_{xh_b}x_t + W_{h_bh_b}h_{b_{t-1}} + b_{h_b}\right) \end{aligned}$$

h_f and h_b are the hidden layers of the forward layer and the backward layer, respectively. The output disease diagnosis probability is expressed as

$$p\left(y|X\right) = \sigma\left(W_{h_fz}h_t + W_{h_bz}h_t + b_z\right)$$

where W_{h_fz}, W_{h_bz} and b_z are the parameters to be trained. In order to prevent overfitting of the training model, dropout is added to the acyclic part of BLSTM. Optimize the cross entropy loss function L_θ by back propagation through time (BPTT).

$$L_\theta = \frac{1}{n}\sum_{i=1}^{n} y \ln\left(p\left(y|X_i\right)\right) + \left(1 - y\right)\ln\left(1 - p\left(y|X_i\right)\right)$$

IV. EXPERIMENT AND ANALYSIS

A. EXPERIMENTAL DATA AND SETTINGS

The experimental data of this study are self-made data, which are thyroid disease, diabetes, hypertension, and coronary heart disease obtained from the database of a hospital. 1200 cases of each disease were selected. For the diagnosis of thyroid disease, the other three types of diseases are all negative samples. The details of the data set are shown in **TABLE 2**.

TABLE 2. Description of experimental data set.

Total number of data sets		4800	
Positive samples	1200	Negative samples	3600
Training sets	3600	Test sets	1200

TABLE 3. Evaluation index.

Index	Calculation formula
Accuracy	$\frac{TP+TN}{TP+FP+TN+FN}$
Precision	$\frac{TP}{TP+FP}$
Recall	$\frac{TP}{TP+FN}$
F_1	$2\times\frac{\text{Precision}}{\text{Precision}+\text{Recall}}$
Remarks	TP and TN respectively represent the number of positive samples and negative samples under the correct classification., FP and FN respectively represent the number of negative samples classified as positive samples and the number of positive samples classified as negative samples.

In order to verify the feasibility and effectiveness of the proposed thyroid disease diagnosis model, comparison diagnosis models include SVM [43], BPNN [44], RNN [45], LSTM [46]. The evaluation indicators are Accuracy, Precision, Recall and F1. The calculation formula of each index is shown in **TABLE 3.** For the evaluation of two classifications, the classification accuracy rate cannot be considered alone, and the four evaluation indicators in Table 4 should be considered comprehensively.

B. EXPERIMENTAL RESULTS AND ANALYSIS

1) COMPARISON OF DIAGNOSIS RESULTS BASED ON DIFFERENT FEATURE EXTRACTION METHODS

In order to explore the impact of different feature extraction methods on the accuracy of disease diagnosis, three comparison experiments were set up. The first group uses LDA to generate structured knowledge, which is the topic. The

TABLE 4. Performance comparison of different feature extraction methods (%).

Index\Method	Knowledge Graph	LDA	VSM
Accuracy	83.24	70.03	74.25
Precision	83.18	66.80	72.49
Recall	81.15	70.81	78.85
F1	81.76	68.72	77.75

TABLE 5. Comparison of disease diagnosis performance of different classifiers (%).

Index\Method	SVM	BPNN	RNN	LSTM	Our method
Accuracy	83.24	81.36	85.68	86.19	88.86
Precision	83.18	80.52	84.27	86.02	87.45
Recall	81.15	78.90	82.62	83.66	89.82
F1	81.76	78.95	81.89	83.10	89.40

number of topics for the thyroid gland is set to be the same as the number of structured features we extracted for the disease. The second group uses the space vector model (VSM) to represent the text. The third group is the method of medical knowledge graph mentioned in this article. The classifier used is the classic SVM. The five-fold crossover method was used to verify the performance of each method. The experimental results are shown in **TABLE 4** and **FIGURE 8**.

The results show that the classification results based on the structured features extracted from the knowledge graph are better than those based on the vector space model. The classification result based on LDA feature extraction is the worst. It shows that the proposed structured features obtained by using the knowledge map have advantages in disease recognition tasks. In order to ensure the robustness of the experiment, multiple identical experiments with multiple diseases have been performed. The experimental results of our method are stable, and the recognition rate of thyroid diseases is above 80%.

2) COMPARISON OF DIAGNOSIS RESULTS BASED ON DIFFERENT CLASSIFIERS

From the experimental results in the previous section, it can be observed that the classification performance based on the knowledge map features is the best. In order to explore the impact of different classifiers on disease diagnosis results, this study introduces multiple contrast classifiers. The feature extraction method of knowledge graph is adopted here. Five-fold cross-validation was used to verify the performance of each method. The experimental results are shown in **TABLE 5** and **FIGURE 9**.

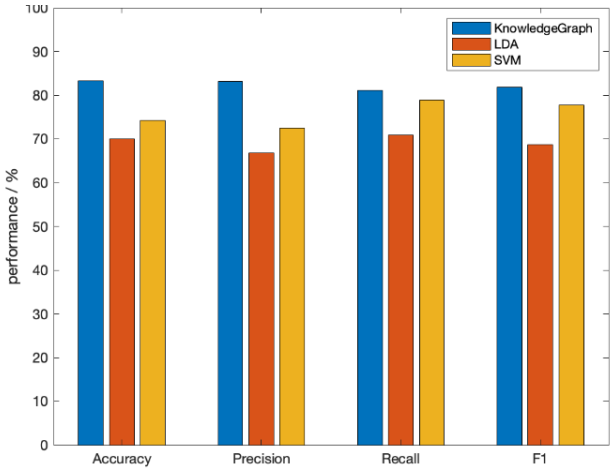

FIGURE 8. Comparison of 4 evaluation indexes of each feature extraction method.

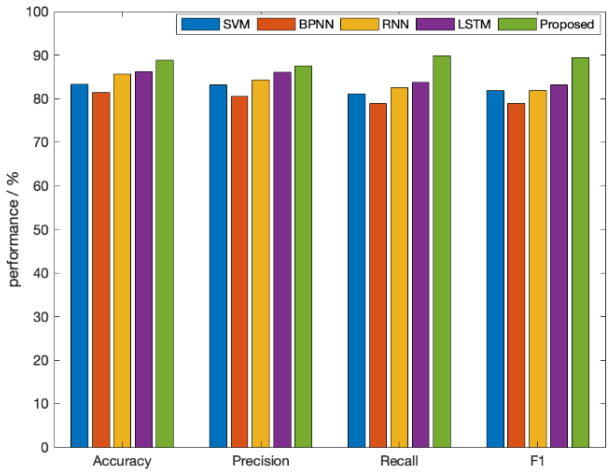

FIGURE 9. Comparison of disease diagnosis performance of different classifiers.

From the experimental results, it can be observed that the four index data of SVM and BPNN algorithm are lower than the four index data obtained by RNN, LSTM and our method. This shows that deep learning algorithms are better than machine learning algorithms for classification. Among the three deep learning algorithms, our method has a leading advantage in four indicators, followed by LSTM, and RNN is the worst. This is because the BLSTM model can better capture bidirectional semantics. In terms of network structure, because it considers both forward LSTM and backward LSTM, it is more robust.

V. CONCLUSION

As people pay more attention to health, many netizens will consult the disease through the Internet, and a large amount of disease description texts have been produced. In addition, the hospital's information platform stores a large amount of disease, disease and other information. How to effectively

use this information for intelligent diagnosis of diseases is the ultimate goal of this research. Since most illnesses are recorded using texts, in view of the diversity of the recorded texts of illnesses, this research proposes a diagnosis method for thyroid diseases based on knowledge graphs and deep learning. This method first extracts the relationships between entities in the biomedical literature to construct a biomedical knowledge graph. Then, the entities and relationships in the knowledge graph are transformed into low-dimensional continuous vectors through the knowledge graph embedding method. Finally, the known pathological disease relationship data is used to train the disease diagnosis model based on BLSTM. Experiments verify the effectiveness of our method in the diagnosis of thyroid diseases. There are two innovations in this research. One is to use knowledge graphs to extract structured features and complete pathological structured text representation. The structured representation based on the knowledge graph is a new structured knowledge extraction method, which can be used for the structured knowledge extraction of disease conditions. The second point is the use of classifiers based on deep learning algorithms. Compared with traditional machine learning algorithms, it greatly improves the accuracy of disease diagnosis.

REFERENCES

[1] W. Lie, B. Jiang, and W. Zhao, "Obstetric imaging diagnostic platform based on cloud computing technology under the background of smart medical big data and deep learning," *IEEE Access*, vol. 8, pp. 78265–78278, 2020.

[2] S. H. Ghasemi, K. Etminani, H. Dehghan, S. Eslami, M. R. Hasibian, H. Vakili-Arki, M. R. Saberi, M. Aghabagheri, and S. M. Namayandeh, "Design and evaluation of a smart medication recommendation system for the electronic prescription," *Studies Health Technol. Inform.*, vol. 260, pp. 128–135, Jun. 2019.

[3] T. Goodwin and S. M. Harabagiu, "Automatic generation of a qualified medical knowledge graph and its usage for retrieving patient cohorts from electronic medical records," in *Proc. 7th IEEE Int. Conf. Semant. Comput.*, Irvine, CA, USA, 2013, pp. 363–370.

[4] X. Zhang, Q. Yang, J. Ding, and Z. Wang, "Entity profiling in knowledge graphs," *IEEE Access*, vol. 8, pp. 27257–27266, 2020.

[5] N. Guan, D. Song, and L. Liao, "Knowledge graph embedding with concepts," *Knowl.-Based Syst.*, vol. 164, pp. 38–44, Jan. 2019.

[6] F. M. Suchanek, G. Kasneci, and G. Weikum, "Yago: A core of semantic knowledge," in *Proc. 16th Int. Conf. World Wide Web*, New York, NY, USA, May 2007, pp. 697–706.

[7] S. Auer, C. Bizer, G. Kobilarov, J. Lehmann, R. Cyganiak, and Z. Ives, "DBpedia: A nucleus for a Web of open data," in *Proc. Int. Semantic Web Conf.* Berlin, Germany: Springer, 2007.

[8] A. Carlson, J. Betteridge, B. Kisiel, B. Settles, E. R. Hruschka, Jr., and T. M. Mitchell, "Toward an architecture for never-ending language learning," in *Proc. 24th AAAI Conf. Artif. Intell.* Atlanta, GA, USA: AAAI, Jul. 2010, p. 3.

[9] K. Bollacker, C. Evans, P. Paritosh, T. Sturge, and J. Taylor, "Freebase: A collaboratively created graph database for structuring human knowledge," in *Proc. ACM SIGMOD Int. Conf. Manage. Data (SIGMOD)*, New York, NY, USA, 2008, pp. 1247–1250.

[10] P. Chen, Y. Lu, V. W. Zheng, X. Chen, and B. Yang, "KnowEdu: A system to construct knowledge graph for education," *IEEE Access*, vol. 6, pp. 31553–31563, 2018.

[11] V. P. Tel'nov and Y. A. Korovin, "Semantic Web and knowledge graphs as an educational technology of personnel training for nuclear power engineering," *Izvestiya Wysshikh Uchebnykh Zawedeniy, Yadernaya Energetika*, vol. 2019, no. 2, pp. 219–229, Jun. 2019.

[12] T. Pham, X. Tao, J. Zhang, and J. Yong, "Constructing a knowledge-based heterogeneous information graph for medical health status classification," *Health Inf. Sci. Syst.*, vol. 8, no. 1, p. 10, Dec. 2020.

[13] L. Li, P. Wang, J. Yan, Y. Wang, S. Li, J. Jiang, Z. Sun, B. Tang, T.-H. Chang, S. Wang, and Y. Liu, "Real-world data medical knowledge graph: Construction and applications," *Artif. Intell. Med.*, vol. 103, Mar. 2020, Art. no. 101817.

[14] R. S. Ledley and L. B. Lusted, "Reasoning foundations of medical diagnosis," *Science*, vol. 130, no. 3306, pp. 9–21, 1959.

[15] S.-H. Wang, Y. Zhang, Y.-J. Li, W.-J. Jia, F.-Y. Liu, M.-M. Yang, and Y.-D. Zhang, "Single slice based detection for Alzheimer's disease via wavelet entropy and multilayer perceptron trained by biogeography-based optimization," *Multimedia Tools Appl.*, vol. 77, no. 9, pp. 10393–10417, May 2018.

[16] S. Wang, J. Sun, I. Mehmood, C. Pan, Y. Chen, and Y. Zhang, "Cerebral micro-bleeding identification based on a nine-layer convolutional neural network with stochastic pooling," *Concurrency Comput., Pract. Exper.*, vol. 32, no. 1, Jan. 2020, Art. no. e5130.

[17] Y.-D. Zhang, V. V. Govindaraj, C. Tang, W. Zhu, and J. Sun, "High performance multiple sclerosis classification by data augmentation and AlexNet transfer learning model," *J. Med. Imag. Health Informat.*, vol. 9, no. 9, pp. 2012–2021, Dec. 2019.

[18] Y. Zhang, S. Wang, Y. Sui, M. Yang, B. Liu, H. Cheng, J. Sun, W. Jia, P. Phillips, and J. M. Gorriz, "Multivariate approach for Alzheimer's disease detection using stationary wavelet entropy and predator-prey particle swarm optimization," *J. Alzheimer's Disease*, vol. 65, no. 3, pp. 855–869, Sep. 2018.

[19] C. Kang, X. Yu, S.-H. Wang, D. Guttery, H. Pandey, Y. Tian, and Y. Zhang, "A heuristic neural network structure relying on fuzzy logic for images scoring," *IEEE Trans. Fuzzy Syst.*, early access, Jan. 13, 2020, doi: 10.1109/TFUZZ.2020.2966163.

[20] S.-H. Wang, Y.-D. Zhang, M. Yang, B. Liu, J. Ramirez, and J. M. Gorriz, "Unilateral sensorineural hearing loss identification based on double-density dual-tree complex wavelet transform and multinomial logistic regression," *Integr. Comput.-Aided Eng.*, vol. 26, no. 4, pp. 411–426, Sep. 2019.

[21] S.-H. Wang, J. Sun, P. Phillips, G. Zhao, and Y.-D. Zhang, "Polarimetric synthetic aperture radar image segmentation by convolutional neural network using graphical processing units," *J. Real-Time Image Process.*, vol. 15, no. 3, pp. 631–642, Oct. 2018.

[22] S. Wang, C. Tang, J. Sun, and Y. Zhang, "Cerebral micro-bleeding detection based on densely connected neural network," *Frontiers Neurosci.*, vol. 13, May 2019, Art. no. 422.

[23] S.-H. Wang, S. Xie, X. Chen, D. S. Guttery, C. Tang, J. Sun, and Y.-D. Zhang, "Alcoholism identification based on an AlexNet transfer learning model," *Frontiers Psychiatry*, vol. 10, Apr. 2019, Art. no. 205.

[24] E. H. Shortliffe, S. G. Axline, B. G. Buchanan, T. C. Merigan, and S. N. Cohen, "An artificial intelligence program to advise physicians regarding antimicrobial therapy," *Comput. Biomed. Res.*, vol. 6, no. 6, pp. 544–560, Dec. 1973.

[25] D. A. Heckerman, "A tractable inference algorithm for diagnosing multiple diseases," *Mach. Intell. Pattern Recognit.*, vol. 10, pp. 163–171, Jan. 1990.

[26] K. Baati, T. M. Hamdani, and A. M. Alimi, "Diagnosis of lymphatic diseases using a naive bayes style possibilistic classifier," in *Proc. IEEE Int. Conf. Syst., Man, Cybern.*, vol. 8215, no. 2, Oct. 2013, pp. 4539–4542.

[27] A. F. Otoom, E. E. Abdallah, and Y. Kilani, "Effective diagnosis and monitoring of heart disease," *Int. J. Softw. Eng. Appl.*, vol. 9, no. 1, pp. 143–156, 2015.

[28] S. Vijayarani and S. Dhayanand, "Liver disease prediction using SVM and Naïve Bayes algorithms," *Eng. Technol. Res.*, vol. 4, pp. 816–820, Apr. 2015.

[29] V. Kunwar, K. Chandel, A. S. Sabitha, and A. Bansal, "Chronic Kidney Disease analysis using data mining classification techniques," in *Proc. 6th Int. Conf.*, Noida, India, Jan. 2016, pp. 300–305.

[30] R. A. Miller, H. E. Pople, and J. D. Myers, "Internist-I, an experimental computer-based diagnostic consultant for general internal medicine," *New England J. Med.*, vol. 307, no. 8, pp. 468–476, Aug. 1982.

[31] H. L. Semigran, D. M. Levine, S. Nundy, and A. Mehrotra, "Comparison of physician and computer diagnostic accuracy," *JAMA Internal Med.*, vol. 176, no. 12, pp. 1860–1861, 2016.

[32] H. Gao, W. Huang, and X. Yang, "Applying probabilistic model checking to path planning in an intelligent transportation system using mobility trajectories and their statistical data," *Intell. Autom. Soft Comput.*, vol. 25, no. 3, pp. 547–559, 2019.

[33] J. Yu, Z. Kuang, B. Zhang, W. Zhang, D. Lin, and J. Fan, "Leveraging content sensitiveness and user trustworthiness to recommend fine-grained privacy settings for social image sharing," *IEEE Trans. Inf. Forensics Security*, vol. 13, no. 5, pp. 1317–1332, May 2018.
[34] H. Gao, W. Huang, Y. Duan, X. Yang, and Q. Zou, "Research on cost-driven services composition in an uncertain environment," *J. Internet Technol.*, vol. 20, no. 3, pp. 755–769, 2019.
[35] J. Yu, C. Hong, Y. Rui, and D. Tao, "Multitask autoencoder model for recovering human poses," *IEEE Trans. Ind. Electron.*, vol. 65, no. 6, pp. 5060–5068, Jun. 2018.
[36] Y. Jiang, K. Zhao, K. Xia, J. Xue, L. Zhou, Y. Ding, and P. Qian, "A novel distributed multitask fuzzy clustering algorithm for automatic MR brain image segmentation," *J. Med. Syst.*, vol. 43, no. 5, pp. 118:1–118:9, May 2019.
[37] P. Qian, J. Zhou, Y. Jiang, F. Liang, K. Zhao, S. Wang, K.-H. Su, and R. F. Muzic, "Multi-view maximum entropy clustering by jointly leveraging inter-view collaborations and intra-view-weighted attributes," *IEEE Access*, vol. 6, pp. 28594–28610, 2018.
[38] Y. Jiang, Z. Deng, F.-L. Chung, G. Wang, P. Qian, K.-S. Choi, and S. Wang, "Recognition of epileptic EEG signals using a novel multiview TSK fuzzy system," *IEEE Trans. Fuzzy Syst.*, vol. 25, no. 1, pp. 3–20, Feb. 2017.
[39] Y. Jiang, D. Wu, Z. Deng, P. Qian, J. Wang, G. Wang, F.-L. Chung, K.-S. Choi, and S. Wang, "Seizure classification from EEG signals using transfer learning, semi-supervised learning and TSK fuzzy system," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 25, no. 12, pp. 2270–2284, Dec. 2017.
[40] P. Qian, Y. Chen, J.-W. Kuo, Y.-D. Zhang, Y. Jiang, K. Zhao, R. Al Helo, H. Friel, A. Baydoun, F. Zhou, J. U. Heo, N. Avril, K. Herrmann, R. Ellis, B. Traughber, R. S. Jones, S. Wang, K.-H. Su, and R. F. Muzic, "MDixon-based synthetic CT generation for PET attenuation correction on abdomen and pelvis jointly using transfer fuzzy clustering and active learning-based classification," *IEEE Trans. Med. Imag.*, vol. 39, no. 4, pp. 819–832, Apr. 2020.
[41] D. Das Chakladar, S. Dey, P. P. Roy, and D. P. Dogra, "EEG-based mental workload estimation using deep BLSTM-LSTM network and evolutionary algorithm," *Biomed. Signal Process. Control*, vol. 60, Jul. 2020, Art. no. 101989.
[42] G. Lopes Andrade and D. Hoffmann Thomas, "An optimized breadth-first search algorithm for routing in optical access networks," *IEEE Latin Amer. Trans.*, vol. 17, no. 07, pp. 1088–1095, Jul. 2019.
[43] N. N. Kulkarni and V. K. Bairagi, "Extracting salient features for EEG-based diagnosis of Alzheimer's disease using support vector machine classifier," *IETE J. Res.*, vol. 63, no. 1, pp. 11–22, Jan. 2017.
[44] H. Guo, W. Zeng, Y. Shi, J. Deng, and L. Zhao, "Kernel granger causality based on back propagation neural network fuzzy inference system on fMRI data," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 28, no. 5, pp. 1049–1058, May 2020.
[45] R. Collobert, J. Weston, L. Bottou, M. Karlen, K. Kavukcuoglu, and P. Kuksa, "Natural language processing (almost) from scratch," *J. Mach. Learn. Res.*, vol. 12 pp. 2493–2537, Aug. 2011.
[46] X. Hong, R. Lin, C. Yang, N. Zeng, C. Cai, J. Gou, and J. Yang, "Predicting Alzheimer's disease using LSTM," *IEEE Access*, vol. 7, pp. 80893–80901, 2019.

XUQING CHAI is currently a Lecturer/Engineer with the College of Computer and Information Engineering, Henan Normal University. She has published more than ten academic articles in national/international journals. Her research interests include computer networks, social networks, cloud computing, and their applications.