KNOWLEDGE-GUIDED MULTIMODAL LARGE MODEL-BASED MEDICAL IMAGE RECOGNITION METHOD

Technical Field

The invention relates to the technical field of medical image analysis, specifically to a knowledge-guided multimodal large model-based medical image recognition method.

Background Art

Medical images (such as X-ray, CT, MRI, etc.) are important bases for clinical diagnosis. In recent years, deep learning-based medical image analysis technology has achieved significant progress, assisting doctors in tasks such as lesion detection and disease classification.

However, most existing technologies rely solely on the image data itself. In use, pure image recognition models lack a deep clinical semantic understanding of the image content.

For example, a model may identify a shadow in the lung but struggle to associate this finding with rich clinical knowledge such as specific diseases, their typical symptoms, and common locations. This leads to poor clinical interpretability of the recognition results and is prone to misjudgments that violate medical common sense.

Furthermore, in actual clinical work, imaging examinations are typically accompanied by rich textual information, such as imaging reports, clinical history, and laboratory test results. Existing methods fail to effectively and deeply integrate these non-image, valuable prior knowledge textual data with image data, resulting in information silos.

Currently, although some multimodal learning methods have emerged attempting to combine images and text, they usually only perform simple fusion at the data level, lacking explicit utilization of structured medical knowledge.

Summary of the Invention

The technical problem to be solved by the invention is to overcome the hereinabove technical shortcomings and provide a knowledge-guided multimodal large model-based medical image recognition method that enhances the model's semantic understanding capability for medical images and improves the accuracy and clinical interpretability of

recognition results.

To solve the hereinabove technical problem, the technical solution provided by the invention is a knowledge-guided multimodal large model-based medical image recognition method, comprising the following steps:

S1: acquiring medical image data to be recognized and non-image clinical text data related to the medical image;

S2: constructing a multimodal large model;

S3: inputting the multi-source data collected in S1 into the multimodal large model obtained in S2 to perform joint encoding, generating a multimodal representation that fuses image semantics and text semantics;

S4: extracting structured medical knowledge information related to the multimodal representation according to a medical knowledge graph;

S5: using the structured medical knowledge information to perform knowledge-guided adjustment on the multimodal representation, enhancing semantic understanding of lesion areas in the medical image;

S6: outputting a recognition result of the medical image based on the guided multimodal representation.

Preferably, the medical image data in S1 comprises at least one of X-ray, CT, MRI, ultrasound, or pathological slide images;

the non-image clinical text data comprises text information from clinical reports, diagnostic conclusions, medical history summaries, or radiology examination reports.

Preferably, the multimodal large model constructed in step S2 comprises an image encoder, a text encoder, and a multimodal fusion module;

the image encoder is configured to perform visual semantic extraction on the input medical image data, generating an image representation vector;

the text encoder is configured to perform semantic parsing on the input non-image clinical text data based on a pre-trained language model, generating a text representation vector;

the multimodal fusion module is configured to perform cross-modal alignment and semantic interaction on the image representation vector and the text representation vector,

generating a joint multimodal representation that fuses image and text semantic information.

Preferably, S4 further comprises:

S4.1: parsing the multimodal representation generated in step S3 to extract key medical entities and relationships;

S4.2: using the key medical entities and relationships as query conditions to search and match within the medical knowledge graph, acquiring related diseases, symptoms, anatomical structures, pathological features, and their interrelationships, thereby forming said structured medical knowledge information.

Preferably, the medical knowledge graph comprises at least one type of structured knowledge among disease-symptom associations, disease-imaging manifestation associations, anatomical structure-lesion relationships, or clinical guideline rules.

Preferably, the step S5 comprises introducing a knowledge-driven attention mechanism to transform the structured medical knowledge information into learnable prompt vectors or knowledge embeddings.

Preferably, the construction of the multimodal large model in S2 comprises using annotated medical image-text pair data during the training phase and performing joint optimization in combination with triple relationships in the medical knowledge graph.

Preferably, the recognition result in S6 comprises at least one of lesion type, location, and clinical relevance assessment.

Preferably, the medical knowledge graph is a dynamically updated knowledge base, including online updates and manual updates.

Compared with the prior art, the advantages of the invention are as follows:

by introducing a medical knowledge graph as prior knowledge, errors that may occur in purely data-driven models and violate medical common sense can be effectively corrected, making the model's reasoning process more rational, thereby improving recognition accuracy for complex and rare cases and the model's robustness;

the invention fully utilizes image and text data that are readily available in clinical practice, reducing the model's dependence on a large amount of finely annotated pixel-level data. It not only achieves the fusion of representations from image and text modalities but further achieves deep integration between the clinical knowledge level and data features

through the knowledge graph, endowing the model with certain medical reasoning capabilities, making it easy to promote and use;

Brief Description of the Drawings

Fig. 1 is a schematic structural diagram of a knowledge-guided multimodal large model-based medical image recognition method.

Specific Embodiment of the invention

A detailed description of the invention is provided hereinafter in conjunction with the drawings.

As shown in conjunction with Fig. 1, the knowledge-guided multimodal large model-based medical image recognition method comprises the following steps: S1: acquiring medical image data to be recognized and non-image clinical text data related to the medical image; S2: constructing a multimodal large model;

S3: inputting the multi-source data collected in S1 into the multimodal large model obtained in S2 to perform joint encoding, generating a multimodal representation that fuses image semantics and text semantics; S4: extracting structured medical knowledge information related to the multimodal representation according to a medical knowledge graph;

S5: using the structured medical knowledge information to perform knowledge-guided adjustment on the multimodal representation, enhancing semantic understanding of lesion areas in the medical image; S6: outputting a recognition result of the medical image based on the guided multimodal representation.

In one embodiment, the medical image data in S1 comprises at least one of X-ray, CT, MRI, ultrasound, or pathological slide images;

the non-image clinical text data comprises text information from clinical reports, diagnostic conclusions, medical history summaries, or radiology examination reports, adapting video and text information for multi-source data acquisition and analysis. The multimodal large model constructed in S2 comprises an image encoder, a text encoder, and a multimodal fusion module;

the image encoder is configured to perform visual semantic extraction on the input medical image data, generating an image representation vector;

the text encoder is configured to perform semantic parsing on the input non-image clinical text data based on a pre-trained language model, generating a text representation vector;

the multimodal fusion module is configured to perform cross-modal alignment and semantic interaction on the image representation vector and the text representation vector, generating a joint multimodal representation that fuses image and text semantic information.

And S4 further comprises:

S4.1: parsing the multimodal representation generated in step S3 to extract key medical entities and relationships;

S4.2: using the key medical entities and relationships as query conditions to search and match within the medical knowledge graph, acquiring related diseases, symptoms, anatomical structures, pathological features, and their interrelationships, thereby forming said structured medical knowledge information.

In more detail:

the medical knowledge graph comprises at least one type of structured knowledge among disease-symptom associations, disease-imaging manifestation associations, anatomical structure-lesion relationships, or clinical guideline rules;

in one embodiment:

Step S5 comprises introducing a knowledge-driven attention mechanism to transform the structured medical knowledge information into learnable prompt vectors or knowledge embeddings. The construction of the multimodal large model in S2 comprises using annotated medical image-text pair data during the training phase and performing joint optimization in combination with triple relationships in the medical knowledge graph.

The recognition result in S6 comprises at least one of lesion type, location, and clinical relevance assessment. In use, the medical knowledge graph is a dynamically updated knowledge base, including online updates and manual updates.

During specific implementation of the invention,

medical images of a patient (such as X-ray, CT, MRI, etc.) and associated clinical text information (such as medical history, examination reports, diagnostic opinions, etc.) are synchronously acquired from medical imaging devices or information systems. The

multi-source heterogeneous data is input into a pre-constructed and trained multimodal large model. This model is configured to simultaneously understand image content and text semantics, and generate a unified multimodal representation;

the built-in medical knowledge graph is invoked to perform intelligent retrieval of key medical concepts involved in the current case (such as anatomical structures, disease names, typical imaging manifestations, etc.), extracting the structured knowledge most relevant to the current representation. This is introduced into the reasoning process through an attention guidance mechanism, causing the model to focus more on areas or patterns that conform to medical common sense when analyzing the image, thereby suppressing unreasonable or clinically illogical predictions;

the model then outputs comprehensive recognition results including lesion type, spatial location, and possible clinical significance, and optionally generates natural language explanations or visual prompts for doctor reference.

In the invention, the entire process does not rely on a large amount of pixel-level annotation, and the knowledge graph supports regular updates, ensuring the system continuously adapts to the latest medical knowledge and diagnostic and treatment guidelines.

Descriptions of content not detailed in this specification belong to the prior art known to those skilled in the art.

The working principle of the invention is as follows: it organically combines medical images, clinical text, and structured medical knowledge, the multimodal large model performs joint encoding of images and text to obtain a fused semantic representation; subsequently, the medical knowledge graph is used to retrieve prior knowledge related to the current case, such as diseases, anatomical structures, and imaging manifestations; the above knowledge is injected into the model's reasoning process in the form of prompt vectors or attention constraints, guiding it to focus on lesion areas that conform to medical logic.

Ultimately, recognition results combining accuracy and clinical interpretability are output, this mechanism not only enhances the model's ability to discriminate complex or rare lesions but also reduces reliance on a large amount of finely annotated data, making AI reasoning closer to real clinical thinking.

It should be noted that similar reference numerals and letters denote similar items in the

following drawings. Therefore, once an item is defined in one drawing, it does not require further definition and explanation in subsequent drawings.

The hereinabove describes the invention and its embodiments, this description is not limiting, and what is shown in the drawings is merely one embodiment of the invention, the actual structure is not limited to this. In summary, if ordinary technicians in the field are inspired by it and, without departing from the creative purpose of the invention, design structural schemes and embodiments similar to the technical solution without creative effort, they should all fall within the protection scope of the invention.