

Claims

1. A medical imaging diagnostic intelligent agent based on multimodal large model, characterized by comprising:

5 multimodal data access module, used to receive and standardize medical imaging data and associated clinical text data of patients;

multimodal large model analysis engine, connected to the data access module, is used for deep fusion analysis of standardized image and text data to understand cross modal semantic associations;

10 the diagnostic result and evidence generation module is connected to the analysis engine, and is used to generate preliminary diagnosis conclusions and corresponding confidence levels based on the output of the analysis engine, and synchronously generate visual evidence and textual evidence supporting the conclusion.

- 15 2. The medical imaging diagnostic intelligent agent based on multimodal large model according to claim 1, characterized in that: the multimodal large model analysis engine comprises:

visual encoding unit, used to extract hierarchical visual features of medical images;

20 text encoding unit, used to extract semantic features of clinical text;

cross modal fusion unit, used to establish and strengthen the association between visual features and semantic features through attention mechanism, forming a unified cross modal representation.

- 25 3. The medical imaging diagnostic intelligent agent based on multimodal large model

according to claim 2, characterized in that: the visual evidence generated by the diagnostic results and evidence generation module comprises a heatmap highlighting the key areas that contribute the most to diagnostic decision-making on the original medical image.

5

4. The medical imaging diagnostic intelligent agent based on multimodal large model according to claim 3, characterized in that: the text evidence generated by the diagnostic results and evidence generation module comprises a structured list of key clinical indications that affect diagnostic decisions and corresponding relationships with imaging findings.

10

5. The medical imaging diagnostic intelligent agent based on multimodal large model according to claim 4, characterized in that: the clinical text data processed by the multimodal data access module comprises but is not limited to chief complaints, current medical history, laboratory test results, and previous diagnostic reports in electronic medical records.

15

6. A medical imaging diagnosis human-machine collaborative system based on multimodal large model, characterized by comprising:

20 a human-machine collaborative decision-making interaction module, connected to the intelligent agent for communication, is used to receive and display the preliminary diagnostic conclusions and evidence generated by the intelligent agent, and provide an interactive interface for doctors to review, question, correct or supplement information, forming a collaborative diagnostic decision;

25

a feedback driven continuous learning framework is used to collect case data

confirmed or corrected by the human-machine collaborative decision-making interaction module, and use this data to incrementally optimize the multimodal large model analysis engine in the intelligent agent in a secure environment.

- 5 7. The medical imaging diagnosis human-machine collaborative system based on multimodal large model according to claim 6, characterized in that: the interactive interface provided by the human-machine collaborative decision-making interaction module comprises:

10 image interaction unit, allowing doctors to circle, annotate, and attach natural language annotations on displayed medical images;

 diagnostic correction unit, allowing doctors to select, modify, or input new diagnostic conclusions from the diagnostic list provided by the intelligent agent;

 report collaborative generation unit, used to integrate preliminary reports generated by intelligent agents with interactive inputs from doctors, and

15 automatically generate final structured diagnostic reports.

8. The medical imaging diagnosis human-machine collaborative system based on multimodal large model according to claim 7, characterized in that: the feedback driven continuous learning framework comprises:

20 security desensitization and data pool construction unit, used for anonymizing case data after collaborative decision-making and storing it in a local high-quality training data pool;

 offline incremental learning unit, used to periodically use data from the high-quality training data pool to fine tune and train the multimodal large model

25 analysis engine;

model validation and update unit, used to perform performance validation on the fine tuned model and deploy it securely to the production environment after passing the validation.

- 5 9. The medical imaging diagnosis human-machine collaborative system based on multimodal large model according to claim 8, characterized in that: the system also comprises a system security and privacy protection module to ensure encryption, access control, and operation traceability of all medical data during transmission, processing, and storage.

10

10. The medical imaging diagnosis human-machine collaborative system based on multimodal large model according to claim 9, characterized in that: the human-machine collaborative decision-making interaction module also comprises a decision log unit, which is used to fully record the initial output of the intelligent agent, all interaction operation sequences of the doctor, and the final diagnostic conclusion, forming an
- 15 auditable decision trajectory.