Supplementary Methods

MRI Protocol

MRI scans were acquired using a variety of scanners from different institutions. MRI scans were acquired using a variety of scanners across three different institutions. The imaging was primarily conducted using 3.0-T units (Siemens Skyra, Philips Achieva, GE MR 750) and 1.5-T units (Toshiba Medical Systems, Alltech Medical Centauri 1.5T), with the Siemens Skyra and Philips Achieva being predominantly utilized. Contrast-enhanced T1-weighted images were obtained using intravenous gadolinium contrast (0.1 mmol/kg, injection rate 4-5 ml/s). The Siemens Skyra protocol parameters for T2-FLAIR included a flip angle of 150°, slice thickness of 5 mm, TR 6000 ms, TE 81 ms, and TI 2030 ms, while the CE-T1 parameters were flip angle 8°, slice thickness 5 mm, TR 1630 ms, and TE 2.3 ms. The Philips Achieva had slightly different parameters, with T2-FLAIR using a flip angle of 90°, slice thickness of 6 mm, TR 1630 ms, and TE 2.3 ms, and CE-T1 using a flip angle of 80°, slice thickness of 5 mm, TR 130 ms, and TE 2.3 ms.

Input-Output prompt

First, briefly summarize the key factors that differentiate:

1. Circumscribed astrocytic gliomas from diffuse gliomas
2. Low-grade gliomas (WHO grade 2 or 3) from high-grade gliomas (WHO grade 4)

Focus on:

a) The most critical MRI features, including appearance on T1, T2, FLAIR, contrast enhancement, and any other distinctive characteristics

b) The role of patient age and sex in glioma classification and prognosis

c) Any other relevant clinical factors

Now, analyze each patient's case in the provided table using:

1. Patient demographics (age and sex)
2. VASARI (Visually AcceSsible Rembrandt Images) features (F1 to F25, F29/30)
3. T2/FLAIR mismatch sign

As you analyze, consider the relative importance of each factor for diagnosis. Assign your own weighting to these factors based on their diagnostic value for differentiating glioma types and grades, keeping in mind that importance may vary based on the specific combination of features present. For each patient, provide your diagnosis in the following format:

PATIENT [Number]:
Age: [X] years
Sex: [Male/Female]
DIAGNOSIS:
Type: [Diffuse glioma / Circumscribed astrocytic glioma]
Grade (if diffuse): [High-grade (WHO grade 4) / Lower-grade (WHO grade 2 or 3)]

Brief explanation (3-5 key points):

- List the most significant factors for this case (including relevant demographic and VASARI features)
- Explain why these factors are most important in this specific case
- Describe how you weighted these factors in your decision, including how age and sex influenced your assessment

If there's significant uncertainty for any patient, mention it and suggest additional tests or imaging studies that could help refine the diagnosis.

After analyzing all patients, provide a summary of your findings, highlighting:

1. Patterns or notable differences across the cases
2. Any observed correlations between patient demographics and specific glioma types or grades
3. The most consistently influential VASARI features across all cases
4. Any challenges or limitations in the diagnostic process based on the provided information

Please provide the diagnosis for all patients in one response, and make sure to complete the diagnosis for every patient in a single answer!

Knowledge-Enhanced Format

1. Task Description

As a professional neuroradiologist, your task is to diagnose brain gliomas based on given MR imaging features. Follow these steps:

(1) Analyze patient demographics and VASARI features

(2) Determine the tumor type (diffuse glioma or circumscribed astrocytic glioma)

(3) If it's a diffuse glioma, assess its grade (high-grade WHO grade 4 or lower-grade WHO grade 2-3)

(4) Provide diagnostic reasoning, listing the most important diagnostic factors

2. Important VASARI Features and Diagnostic Tendencies

VASARI Feature	Circumscribed astrocytic gliomas	Low-grade diffuse gliomas	High-grade diffuse glioma
F1-Tumor Location	Predominantly cerebellum, optic pathway,temporal lobe	Can occur in any location	Can occur in any location
F2-Side of Lesion Center	Usually unilateral,or midline	Usually unilateral	May be bilateral/central
F4-Enhancement Quality	Marked	Minimal or none	Marked/Avid
F5-Proportion Enhancing	Variable	Usually <33% or none	Variable
F6-Proportion nCET	Variable	Commonly 68-95%	Variable
F7-Proportion Necrosis	Usually <5% or none	Typically none or <5%	common
F8-Cysts	Common (large cysts)	Small cysts (if present)	Small cysts (if present)
F9-Multifocal or Multicentric	Usually unifocal	Usually unifocal	May be multifocal
F10-T1/FLAIR Ratio	Typically expansion (T1=FLAIR)	Variable, Mixed(T1<FLAIR)	May show infiltrative pattern (T1<<FLAIR)
F11-Thickness of enhancing margin	Thick when present	Thin/none for grade 2, Thick for grade 3	Thick, irregular (around necrosis)
F13-Definition of the non-enhancing margin	Well-defined	Poorly defined	Poorly defined
F14-Proportion of Edema	Usually 6-33% or none	Variable	Typically extensive
F15-Edema Crosses Midline	Rare	Possible	May occur
F16-Hemorrhage	Occasionally present	Rare for grade 2, Common for grade 3	Common
F17-Diffusion	Occasionally present	Rare for grade 2, Common for grade 3	Common
F22-nCET tumor Crosses Midline	Rarely	May occur	Common
F23-Enhancing tumor Crosses Midline	Rarely	Rarely	Common
F24-Satellites	none	Uncommon	May occur
F29/30-Lesion Size(mm)	Variable	Variable	Variable，often large

T2/FLAIR mismatch sign： More common in astrocytomas, IDH-mutant

3. Common Glioma Types Characteristics

3.1 Glioblastoma

Most common adult primary brain tumor, aggressive, therapy-resistant, poor prognosis;

Typically heterogeneous masses in white matter with irregular peripheral enhancement, central necrosis, surrounding edema;

Epidemiology: Mainly in adults over 40, Peak incidence 65-75 years

MRI features:

① T1: Hypo- to isointense mass in white matter; central heterogeneous signal (necrosis, intratumoral hemorrhage)

② T1+C: Enhancement almost always present; typically peripheral and irregular with nodular components; usually surrounds necrosis

③ T2/FLAIR: Hyperintense; surrounded by vasogenic edema; flow voids occasionally seen;

④ DWI/ADC: Solid component: Often high DWI signal

3.2 Astrocytoma, IDH-mutant

WHO CNS grade 2, 3 or 4 diffuse infiltrating astrocytic tumors of the brain in adults

No identifiable border between tumor and normal brain tissue, even if borders appear relatively well-marginated on imaging

Epidemiology:

Typically diagnosed in young adults (median age 36 for grades 2 and 3, 38 for grade 4)

Substantially younger than glioblastoma IDH-wildtype tumors (median 50-60 years)

Higher incidence in men of all ages and grades (M:F ~1.5)

MRI features:

① T1: Iso- to hypointense compared to white matter; usually confined to white matter and causes expansion of adjacent cortex

② T2/FLAIR: Mass-like hyperintense signal that incompletely suppresses on FLAIR: T2/FLAIR mismatch sign. Always follow white matter distribution and cause expansion of surrounding cortex

③ Cortex can be involved in late cases (unlike oligodendroglioma, which is cortical-based from the start)

④ "Microcystic changes" along the lines of spread is a unique behavior, but only appreciated in a small number of cases

⑤ High T2 signal is not related to cellularity or cellular atypia, but to edema, demyelination, and other degenerative changes

⑥ DWI/ADC: Typically has facilitated diffusion, with lower ADC values suggesting a higher grade

⑦ T1+C: No enhancement in grade 2 tumors; solid areas of enhancement ± necrosis suggest higher grade

3.3 Oligodendroglioma

Accounts for 5-25% of all gliomas, 5-10% of all primary intracranial neoplasms

Characterized by IDH mutation and 1p19q codeletion, can be WHO CNS grade 2 or 3

Epidemiology:Usually in middle-aged adults, most common in 4th and 5th decades; Grade 3 tumors tend to occur in slightly older patients; Slight male predilection (M:F 1.3:1); Rare in children

MRI features:

① Tumors with intact 1p/19q show more homogeneous signal on T1 and T2 images and have sharper borders

② Well-circumscribed, homogeneously T1 hypointense with high T2 signal and T2/FLAIR mismatch without calcification predicts lack of 1p19q codeletion

③ Minimal peritumoral vasogenic edema in grade 2 tumors

④ T1: Typically hypointense

⑤ T2: Typically hyperintense (except calcified areas)

⑥ T1+C: Contrast enhancement common but not a reliable indicator of tumor grade, with 50% enhancing to a variable degree, usually heterogeneously

⑦ DWI: Typically no diffusion restriction; ADC values can help differentiate from astrocytomas

3.4 Diffuse Midline Glioma, H3 K27-altered

Represents majority of diffuse intrinsic pontine gliomas, also found in other midline locations

Aggressive, poor prognosis, considered WHO grade 4 regardless of histological features

Mainly in young children, located in the pons

Also seen in other midline structures: thalamus, brainstem, spinal cord

MRI features:

① T1: Hypointense

② T2: Hyperintense mass; relative hypointense areas may indicate foci of higher cellularity or spared white matter tracts

③ T1+C: Usually non-enhancing; focal enhancing areas (correlated with relative hypointensity on T2 and low ADC values) may indicate higher-grade areas

④ DWI/ADC: Typically facilitated diffusion; some areas of diffusion restriction may represent higher-grade areas

3.5 Pilocytic Astrocytoma

Often in young patients, primarily in the cerebellum

MRI features:

① Large cystic component with brightly enhancing mural nodule: 67%

② Non-enhancing cyst wall: 21%

③ Enhancing cyst wall: 46%

④ Heterogeneous, mixed solid and multiple cysts and central necrosis: 16%

⑤ Completely solid: 17%

⑥ Enhancement almost always present (~95%)

⑦ Calcification possible (20%)

⑧ T1: Solid component iso- to hypointense; cystic component fluid signal (unless hemorrhagic)

⑨ T1+C: Vivid contrast enhancement; cyst wall enhances in ~50% of cases

⑩ T2: Solid component hyperintense; cystic component high signal

3.6 Pleomorphic Xanthoastrocytoma (PXA)

Common in young patients, can be WHO grade 2 or 3

MRI features:

① Often cortical with cystic component and vivid contrast enhancement, mostly in temporal lobe

② T1: Solid component iso- to hypointense; cystic component hypointense; leptomeningeal involvement in 70% of cases

③ T1+C: Solid component usually enhances vividly

④ T2: Solid component iso- to hyperintense; cystic component hyperintense

⑤ T2/FLAIR: Cystic areas show hyperintensity relative to CSF due to higher protein contents

⑥ Little surrounding vasogenic edema

4. Diagnostic Output Format

Please use the following format for your diagnosis:

```
Patient [Number]:
- Age: [X] years
- Sex: [Male/Female]

Diagnosis:
- Type: [Diffuse glioma / Circumscribed astrocytic glioma]
- Grade (if diffuse): [High-grade (WHO grade 4) / Lower-grade (WHO grade 2 or 3)]

Brief explanation (3-5 key points):
- [List the most significant factors for this case]
- [Explain why these factors are most important in this specific case]
- [Describe how you weighted these factors, including how age and sex influenced your assessment]
```

5. Important Notes

- Consider all available information comprehensively; do not rely solely on a single feature
- For uncertain cases, explain your reasoning process and uncertainties
- If additional information is needed for a more accurate judgment, please state so
- Remember that while you should strive for accurate diagnosis, in actual clinical practice, all features need to be considered comprehensively along with other clinical and pathological information to make accurate diagnoses and prognosis assessments

Error Classification Framework

To systematically analyze diagnostic errors made by LLMs in glioma classification, we established a three-category error taxonomy. Error Type 1 is Diagnostic Reasoning Errors, which refers to failures in logical inference processes during radiological data analysis. Error Type 2 is Feature Processing Errors, which refers to inadequate feature handling, including feature omission where critical VASARI features are excluded from analysis, and feature extraction inconsistency where similar features are interpreted differently across cases. Error Type 3 is Knowledge Application Errors, which refers to inappropriate feature weighting during diagnostic decision-making, where certain features are overemphasized while others are underutilized. For each incorrect classification, three neuroradiologists independently reviewed the LLM's reasoning process to identify and categorize the primary error type, with disagreements resolved through consensus discussion.

Figure S1. LLM performance analysis for diffuse vs. circumscribed glioma classification.

(A-C): ROC curves comparing Input-Output (dashed lines) vs. Knowledge-Enhanced (solid lines) prompting strategies across three LLMs (GPT, Claude3.0-Opus, and Claude3.5-Sonnet) with radiologist input features (R1-R3). (D-F): Corresponding decision curve analysis showing net benefit across threshold probabilities. Knowledge-Enhanced prompting consistently outperformed Input-Output across all models and radiologist inputs, with highest performance achieved by senior radiologist features (green lines) combined with Knowledge-Enhanced prompting, yielding greater net clinical benefit at most decision thresholds.

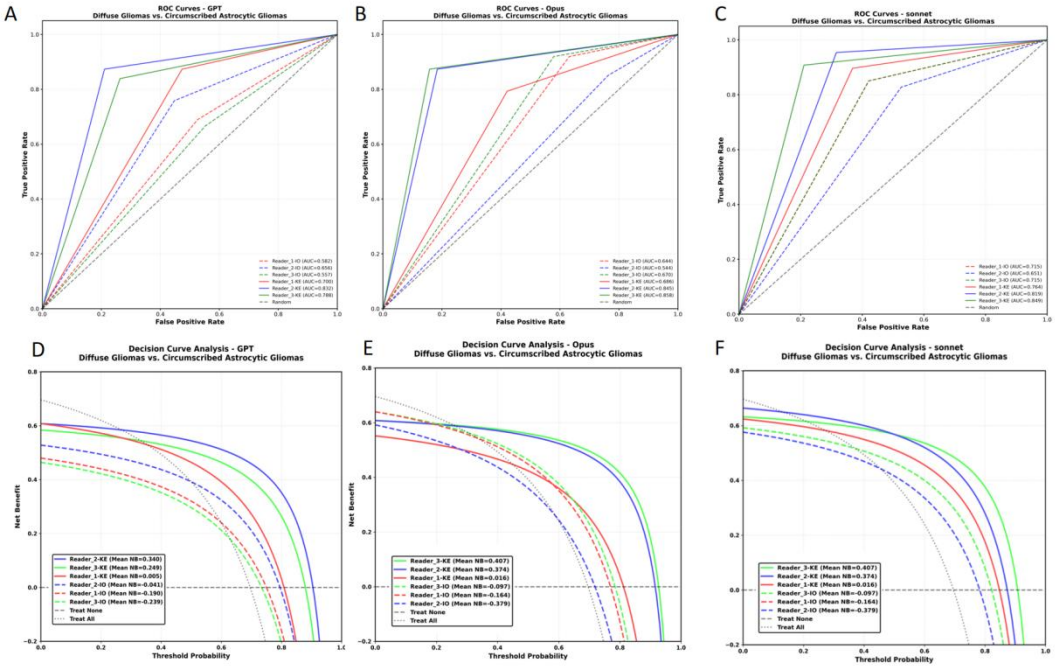

Figure S2. LLM performance analysis for high-grade vs. low-grade glioma classification. (A-C): ROC curves comparing Input-Output vs. Knowledge-Enhanced prompting strategies across three LLMs (GPT, Claude3.0-Opus, and Claude3.5-Sonnet) with radiologist input features (R1-R3). (D-F): Corresponding decision curve analysis showing net benefit across threshold probabilities.

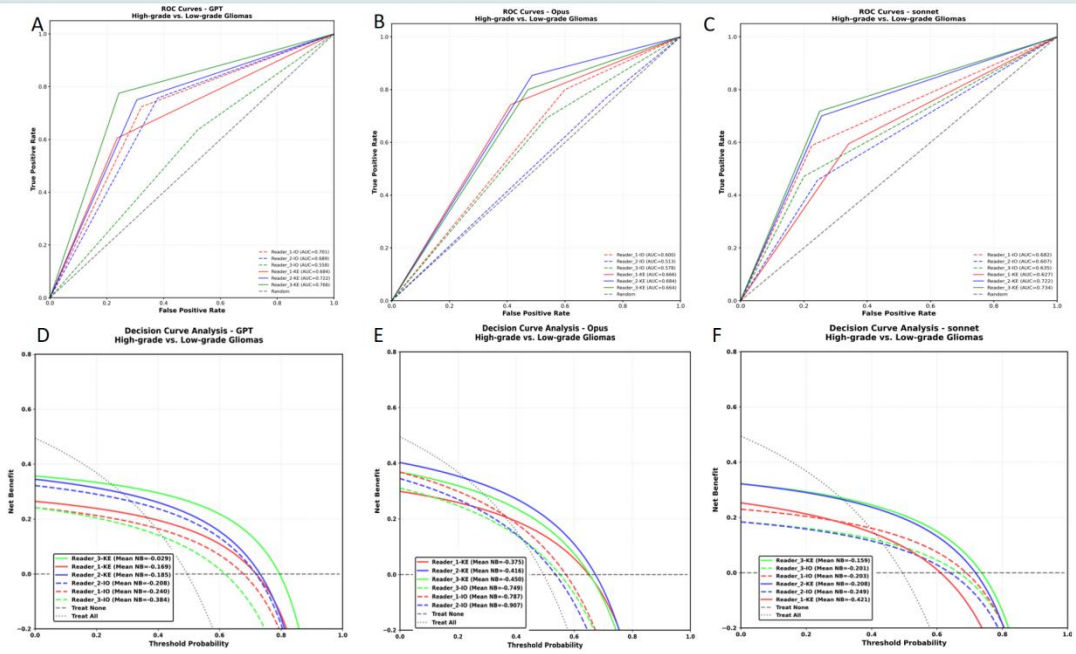

Figure S3. Performance Comparison of Traditional Machine Learning Models and Feature Importance Analysis for Glioma Classification.

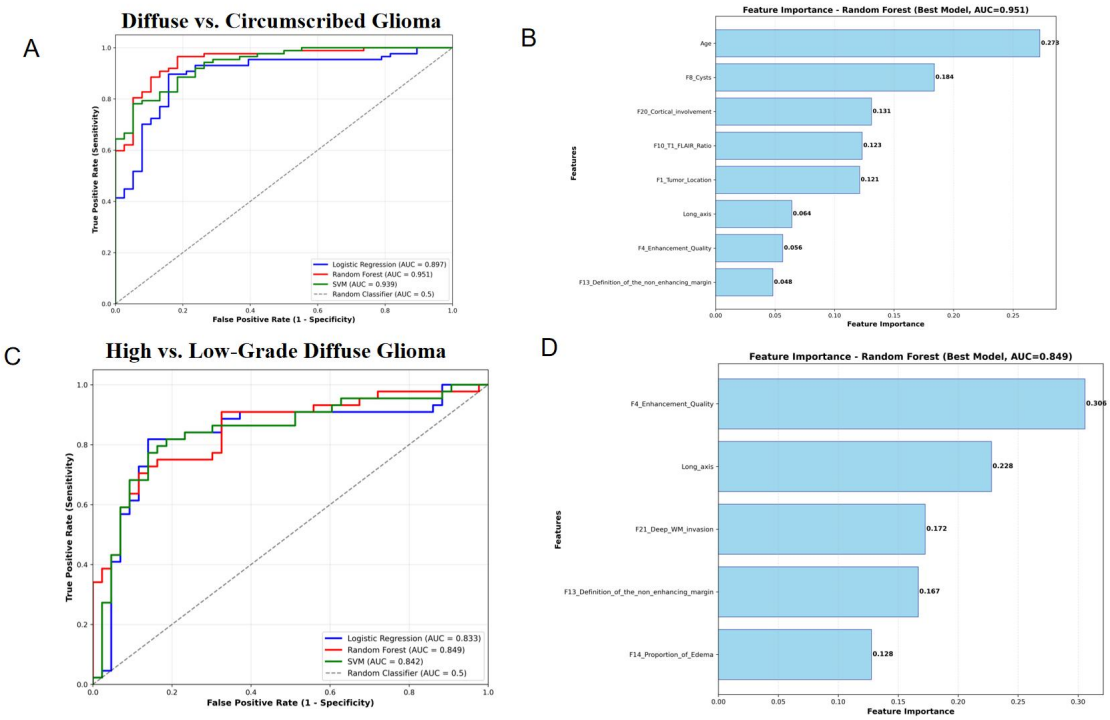

Figure S4. Patient selection flowchart.

Note: CAG = circumscribed astrocytic gliomas; DG-H = diffuse gliomas, high-grade; DG-L = diffuse gliomas, low-grade.

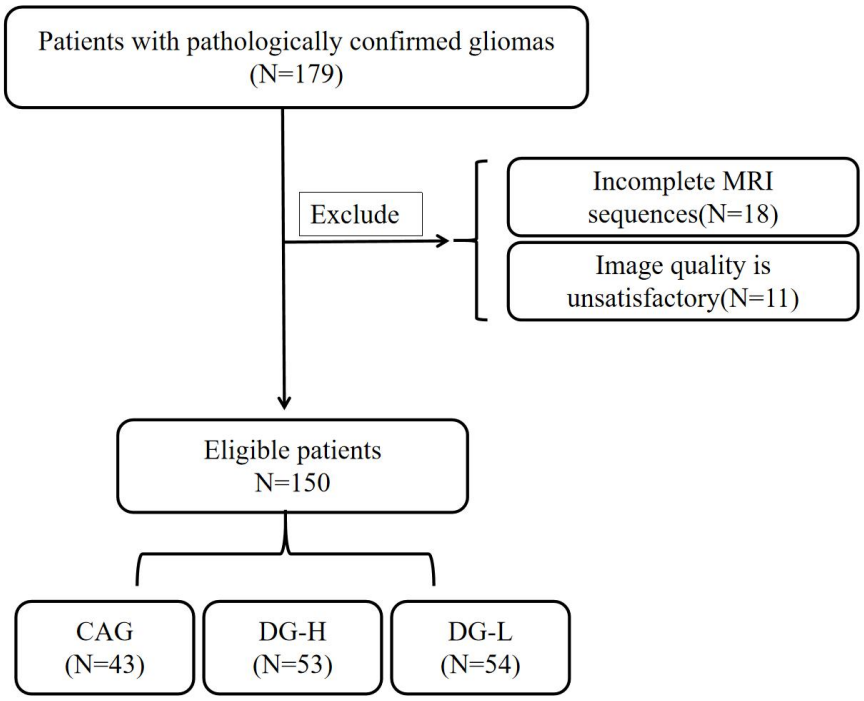

Table S1. Demographic and Clinical Characteristics of patients

Characteristics	Value
Patient number	150
Age (years, mean ± SD)	31 ± 17.44
Gender (male, %)	69, 46%
Diagnosis	
Diffuse Gliomas, high-grade (n, %)	53, 35.3%
Glioblastoma	23
Astrocytoma, IDH-mutant, Grade 4	20
Pediatric diffuse high-grade glioma, H3K27 mutant	10
Diffuse Gliomas, low-grade (n, %)	54, 36.0%
Astrocytoma, IDH-mutant, Grade2-3	32
Oligodendrogliomal, Grade 2-3	22
Circumscribed astrocytic (n, %)	43, 28.7%
Pilocytic astrocytoma	34
Pleomorphic xanthoastrocytoma	9

Table S2. Distribution of Error Types Across LLM Models and Prompting Strategies

	Diagnostic Reasoning Errors (Type 1)	Feature Processing Errors (Type 2)	Knowledge Application Errors (Type 3)
DGs vs. CAGs			
GPT_IO	43.54%	27.46%	29.00%
GPT_KE	48.86%	21.16%	29.98%
Opus_IO	43.50%	20.86%	35.64%
Opus_KE	50.60%	17.65%	31.75%
Sonnet_IO	29.07%	27.80%	43.13%
Sonnet_KE	40.92%	14.09%	44.99%
HG vs. LG			
GPT_IO	39.46%	25.52%	35.02%
GPT_KE	48.08%	11.64%	40.29%
Opus_IO	41.03%	12.05%	46.92%
Opus_KE	47.55%	9.08%	43.37%
Sonnet_IO	33.19%	19.30%	47.52%
Sonnet_KE	42.02%	14.43%	43.55%

Note: DGs = Diffuse Gliomas; CAGs = Circumscribed Astrocytic Gliomas; HG = High-Grade; LG = Low-Grade; IO = Input-Output Format; KE = Knowledge-Enhanced Format.

Table S3. Illustrative Examples of LLM Error Categories

Error Type	Case Background	LLM Error Output	Correct Diagnosis	Error Analysis
Type1: Diagnostic Reasoning Errors	16-year-old male with temporal lobe lesion; LLM analyzed VASARI features: F1-Temporal, F4-Marked enhancement, F5-6-33%, F8-Cysts present, F11-Thick enhancing margin	Age: 16 years Sex: male Diagnosis:Diffuse glioma, Lower-grade (WHO grade 2 or 3) Brief explanation: 1. Key factors: Young age, marked enhancement (F4=3), thickenhancing margin (F11=4), Irregular non-enhancing margin (F13-3) 2. The patient's young age strongly suggests a lower-grade glioma 3. Despite marked enhancement, other features like moderate proportion of enhancement (F5=4, 6-33%) and absence ofmidline crossing support a lower grade 4. Location in temporal lobe (F1=2) with no eloquent brain involvement (F3=1) is typica for diffuse gliomas	Circumscribed astrocytic glioma (Pleomorphic xanthoastrocytoma)	LLM made fundamental diagnostic reasoning errors in its logical inference process: 1. Age-pathology correlation error: Applied incorrect logic that "young age strongly suggests lower-grade diffuse glioma" when young patients (especially adolescents) are actually more likely to develop circumscribed astrocytic tumors. 2. Location-pathology association error: Incorrectly reasoned that "temporal lobe location is typical for diffuse gliomas"
Type2: Feature Processing Errors	Case Background: 6-year-old male with brainstem lesion; LLM analyzed manually assessed VASARI	Age: 6 years Sex:Male Diagnosis: Circumscribed astrocytic glioma likelypilocytic astrocytoma Brief explanation : 1. Key factors: Pediatric	High-grade glioma (Pediatric diffuse high-grade glioma, H3K27 mutant)	LLM committed Feature Processing Errors through significant feature omission in the diagnostic reasoning process: 1. Critical feature omission: LLM completely failed to incorporate key VASARI features F8-yes, F22-yes, and

	features from structured table containing: F1-Brainstem, F4-2 (Mild/minimal), F5-6~33%, F14-<5%, F8-yes, F17-Restricted, F22-yes, F23-yes	patient with infratentorial tumor with cyst, represents typical age and location for circumscribed gliomas such as pilocytic astrocytoma 2. Mild enhancement pattern (F4) affecting only 6-33% of the lesion suggests less aggressive tumor behavior 3. Minimal edema (<5%, F14) indicates limited surrounding tissue disruption, typical of lower-grade or circumscribed lesions 4. Restricted diffusion (F17) shows increased cellularity but can be seen in both low and high-grade gliomas		F23-yes in its diagnostic reasoning, despite these features being present in the structured data and highly indicative of high-grade diffuse pathology. 2. Location feature misprocessing: While LLM identified F1-Brainstem, it misprocessed this critical anatomical feature by treating it as generic "infratentorial location" rather than recognizing brainstem-specific pathology patterns (pilocytic astrocytoma predominantly affects cerebellar hemispheres vs. H3K27 mutant gliomas primarily occur in brainstem and thalamus).
Type3: Knowledge Application Errors	28-year-old male with frontal lobe mass; LLM analyzed manually assessed VASARI features: F1-Frontal, F4-(mark), F5-6-33%, F14-<5%, F17-Restricted, F20-Present, F21-Present	Age: 28 years Sex:Male Diagnosis:Diffuse glioma Grade: Lower-grade (WHO grade 2 or 3) Brief explanation : 1. Patient’s young age (28 years) strongly favors lower-grade pathology, as high-grade gliomas are more common in older patients 2. Mark enhancement pattern (F4) with low proportion (6-33%, F5) suggests less aggressive tumor behavior or anaplastic transformation. 3. Minimal edema (<5%, F14) indicates limited surrounding tissue disruption, characteristic	High-grade glioma (Astrocytoma, IDH-mutant, Grade 4)	LLM demonstrated inappropriate feature weighting allocation during the reasoning process, leading to diagnostic errors: 1. Age overweighting error: LLM overemphasized the age factor (28 years), erroneously believing that young patients inevitably tend toward low-grade pathology, while neglecting that IDH-mutant high-grade gliomas can also occur in young patients. 2. Enhancement pattern misweighting: LLM underestimated the significance of marked enhancement (F4), judging it as low aggressiveness based on moderate enhancement proportion, 3. Aggressive feature

	of lower-grade lesions 4. While restricted diffusion (F17), cortical involvement (F20), and deep white matter invasion (F21) are concerning features that can suggest higher-grade disease, they can also be present in some lower-grade gliomas. 5. Presence of invasion markers (F20, F21) confirms diffuse rather than circumscribed growth pattern, supporting diffuse glioma classification	constellation underweighting: Although LLM correctly identified high-grade indicators including restricted diffusion (F17), cortical involvement (F20), and deep white matter invasion (F21), it erroneously underestimated the importance of these combined features, which were overshadowed by the age factor.