**Supplementary material**
Checklist for the transparent reporting of a multivariable model for individual prognosis or diagnosis for large language models (TRIPOD-LLM).

## The TRIPOD-LLM Statement: A Targeted Guideline For Reporting Large Language Models Use

Fillable TRIPOD-LLM checklist

| Section | Item | Checklist Item | Research Design | LLM Task | Page |
|---|---|---|---|---|---|
| Title | | | | | |
| Title | 1 | Identify the study as developing, fine-tuning, and/or evaluating the performance of an LLM, specifying the task, the target population, and the outcome to be predicted. | All | All | 1,3 |
| Abstract | | | | | |
| Abstract | 2 | See TRIPOD-LLM for Abstracts | All | All | 3 |
| Introduction | | | | | |
| Background | 3a | Explain the healthcare context / use case (e.g., administrative, diagnostic, therapeutic, clinical workflow) and rationale for developing or evaluating the LLM, including references to existing approaches and models. | All | All | 5 |
| | 3b | Describe the target population and the intended use of the LLM in the context of the care pathway, including its intended users in current gold standard practices (e.g., healthcare professionals, patients, public, or administrators). | E<br>H | All | 6 |
| Objectives | 4 | Specify the study objectives, including whether the study describes the initial development, fine-tuning, or validation of an LLM (or multiple stages). | All | All | 6 |
| Methods | | | | | |
| Data | 5a | Describe the sources of data separately for the training, tuning, and/or evaluation datasets and the rationale for using these data (e.g., web corpora, clinical research/trial data, EHR data, or unknown). | All | All | 6,7,8 |
| | 5b | Describe the relevant data points and provide a quantitative and qualitative description of their distribution and other relevant descriptors of the dataset (e.g., source, languages, countries of origin) | All | All | 6, 7 |
| | 5c | Specifically state the date of the oldest and newest item of text used in the development process (training, fine-tuning, reward modeling) and in the evaluation datasets. | All | All | 7,8 |
| | 5d | Describe any data pre-processing and quality checking, including whether this was similar across text corpora, institutions, and relevant socio-demographic groups. | All | All | 7,8 |
| | 5e | Describe how missing and imbalanced data were handled and provide reasons for omitting any data. | All | All | 7 |
| Analytical Methods | 6a | Report the LLM name, version, and last date of training. | All | All | 9 |
| | 6b | Report details of LLM development process, such as LLM architecture, training, fine-tuning procedures, and alignment | M<br>D | All | Not applicable |

| | | | | | |
|---|---|---|---|---|---|
| | | strategy (e.g., reinforcement learning, direct preference optimization, etc.) and alignment goals (e.g., helpfulness, honesty, harmlessness, etc.). | | | |
| | 6c | Report details of how text was generated using the LLM, including any prompt engineering (including consistency of outputs), and inference settings (e.g., seed, temperature, max token length, penalties), as relevant. | M<br>D<br>E | All | 9, 10 |
| | 6d | Specify the initial and post-processed output of the LLM (e.g., probabilities, classification, unstructured text). | All | All | 10,11 |
| | 6e | Provide details and rationale for any classification and, if applicable, how the probabilities were determined and thresholds identified. | All | C<br>OF | 10, 11 |
| LLM Output | 7a | Include metrics that capture the quality of generative outputs, such as consistency, relevance, and accuracy, compared to gold standards. | All | QA<br>IR<br>DG<br>SS<br>MT | Not applicable |
| | 7b | Report the outcome metrics' relevance to downstream task at deployment time and, where applicable, correlation of metric to human evaluation of the text for the intended use. | E<br>H | All | 10 |
| | 7c | Clearly define the outcome, how the LLM predictions were calculated (e.g., formula, code, object, API), the date of inference for closed-source LLMs, and evaluation metrics. | E<br>H | All | 9,10 |
| | 7d | If outcome assessment requires subjective interpretation, describe the qualifications of the assessors, any instructions provided, relevant information on demographics of the assessors, and inter-assessor agreement. | All | All | 10 |
| | 7e | Specify how performance was compared to other LLMs, humans, and other benchmarks or standards. | All | All | 11-13 |
| Annotation | 8a | If annotation was done, report how text was labeled, including providing specific annotation guidelines with examples. | All | All | Not applicable |
| | 8b | If annotation was done, report how many annotators labeled the dataset(s), including the proportion of data in each dataset that were annotated by more than 1 annotator, and the inter-annotator agreement. | All | All | Not applicable |
| | 8c | If annotation was done, provide information on the background and experience of the annotators or characteristics of any models involved in labelling. | All | All | Not applicable |
| Prompting | 9a | If research involved prompting LLMs, provide details on the processes used during prompt design, curation, and selection. | All | All | 10 |
| | 9b | If research involved prompting LLMs, report what data were used to develop the prompts. | All | All | 9, 10 |
| Summarization | 10 | Describe any preprocessing of the data before summarization. | All | SS | Not applicable |
| Instruction tuning/Alig | 11 | If instruction tuning/alignment strategies were used, what were the instructions, data, and interface used for evaluation, | M<br>D | All | Not applicable |

| | | | | | |
|---|---|---|---|---|---|
| nment | | and what were the characteristics of the populations doing evaluation? | | | |
| Compute | 12 | Report compute, or proxies thereof (e.g., time on what and how many machines, cost on what and how many machines, inference time, floating-point operations per second (FLOPs)), required to carry out methods. | M<br>D<br>E | All | Not applicable |
| Ethical Approval | 13 | Name the institutional research board or ethics committee that approved the study and describe the participant-informed consent or the ethics committee waiver of informed consent. | All | All | 6, 7 |
| Open Science | 14a | Give the source of funding and the role of the funders for the present study. | All | All | 2 |
| | 14b | Declare any conflicts of interest and financial disclosures for all authors. | All | All | 21 |
| | 14c | Indicate where the study protocol can be accessed or state that a protocol was not prepared. | H | All | Not applicable |
| | 14d | Provide registration information for the study, including register name and registration number, or state that the study was not registered. | H | All | Not applicable |
| | 14e | Provide details of the availability of the study data. | All | All | 21 |
| | 14f | Provide details of the availability of the code to reproduce the study results. | All | All | 21 |
| Public Involvement | 15 | Provide details of any patient and public involvement during the design, conduct, reporting, interpretation, or dissemination of the study or state no involvement. | H | All | Not applicable |
| Results | | | | | |
| Participants | 16a | When using patient/EHR data, describe the flow of text/EHR/patient data through the study, including the number of documents/questions/participants with and without the outcome/label and follow-up time as applicable. | E<br>H | All | 7, 8 |
| | 16b | When using patient/EHR data, report the characteristics overall and, for each data source or setting, and for development/evaluation splits, including the key dates, key characteristics, and sample size. | E<br>H | All | 12 |
| | 16c | For LLM evaluation that include clinical outcomes, show a comparison of the distribution of important clinical variables that may be associated with the outcome between development and evaluation data, if available. | E<br>H | All | 11-13 |
| | 16d | When using patient/EHR data, specify the number of participants and outcome events in each analysis (e.g., for LLM development, hyperparameter tuning, LLM evaluation). | E<br>H | All | 11-13 |
| Performance | 17 | Report LLM performance according to pre-specified metrics (see item 7a) and/or human evaluation (see item 7d). | All | All | 12,13 |
| LLM Updating | 18 | If applicable, report the results from any LLM updating, including the updated LLM and subsequent performance. | All | All | Not applicable |

| | | | | | |
|---|---|---|---|---|---|
| Discussion | | | | | |
| Interpretation | 19a | Give an overall interpretation of the main results, including issues of fairness in the context of the objectives and previous studies. | All | All | 14 |
| Limitations | 19b | Discuss any limitations of the study and their effects on any biases, statistical uncertainty, and generalizability. | All | All | 16 |
| Usability of the LLM in context | 19c | Describe any known challenges in using data for the specified task and domain context with reference to representation, missingness, harmonization, and bias. | E<br>H | All | 5, 6, 16 |
| | 19d | Define the intended use for the implementation under evaluation, including the intended input, end-user, level of autonomy/human oversight. | E<br>H | All | 13 |
| | 19e | If applicable, describe how poor quality or unavailable input data should be assessed and handled when implementing the LLM, i.e., what is the usability of the LLM in the context of current clinical care. | E<br>H | All | Not applicable |
| | 19f | If applicable, specify whether users will be required to interact in the handling of the input data or use of the LLM, and what level of expertise is required of users. | E<br>H | All | 13 |
| | 19g | Discuss any next steps for future research, with a specific view to applicability and generalizability of the LLM. | All | All | 16 |

LLM = large language model; M = LLM methods; D = *de novo* LLM development; E = LLM evaluation; H = LLM evaluation in healthcare settings; C = classification; OF = outcome forecasting; QA = long-form question-answering; IR = information retrieval; DG = document generation; SS = summarization and simplification; MT = machine translation; EHR = electronic health record.

Note: For studies using existing LLMs, users should include reference(s) to reportable information if provided by the original developers or state that this information is not available.

**Supplementary material**

**The full prompt template and representative input-output examples.**

**Representative case 1:**

**Prompt :** You are an expert in digestive endoscopy evaluating a gastric lesion that was biopsied as low-grade intraepithelial neoplasia (LGIN) before resection. Pathologic upgrade is defined as post-resection histology showing high-grade intraepithelial neoplasia (HGIN) or early gastric cancer. Based on the endoscopic features below—and your knowledge of the literature—please estimate the probability of pathologic upgrade. Please respond in English.
Lesion features are as follows:
Size: 2 cm
Location: gastric body
Background atrophy: moderate
Intestinal metaplasia: present
Paris classification: IIc
Morphology: slightly depressed
Surface ulceration: present
Surface nodularity: absent
Surface redness or erythema: present
Microsurface (MS) pattern: absent
Microvascular (MV) pattern: irregular

**DeepSeek-V3:**

### **High-Risk Lesion: Estimated Upgrade Probability 70-85%**
#### **Critical Risk Features:**
1. **Size (2 cm)**
    - Meets high-risk threshold [ID:0][ID:4]
    - 55-65% upgrade rate in lesions ≥2 cm [ID:6]
2. **Paris IIc Depressed Morphology + Ulceration**
    - Highest-risk macroscopic type (OR 4.7) [ID:4]
    - Ulceration: OR=14.4 (strongest predictor) [ID:3]
    - Present in 83% of malignant upgrades [ID:6]
3. **Microvascular Irregularity**
    - MV irregularity: OR=8.5 [ID:3]
    - Absent MS pattern suggests advanced disruption [ID:5]
4. **Supporting Factors**
    - Erythema present (mild risk) [ID:0]
    - Intestinal metaplasia: OR=4.3 [ID:3]
#### **Probability Estimate:**

- **Baseline LGIN upgrade**: 40.8% [ID:0]
- **This lesion**: Meets 5/6 major risk criteria
- **Final estimate**: **70-85% probability** of HGIN/EGC
#### **Immediate Management Required:**
1. **Therapeutic ESD Mandatory**
    - En bloc resection with 5mm margins [ID:5][ID:6]
    - Hybrid knife recommended for fibrotic areas [ID:7]
2. **Pre-Resection Staging**
    - EUS: 30% submucosal invasion risk [ID:7]
    - CT abdomen if deep invasion suspected [ID:6]
3. **Concurrent Interventions**
    - Gastric mapping (12 biopsies) [ID:5]
    - Mandatory HP eradication [ID:4]

*"This constellation represents endoscopically obvious early gastric cancer requiring immediate curative-intent resection"* [ID:5][ID:7]
Would you like the specific ESD protocol for body lesions?

**Qwen3-235b:**

<think>Okay, let's tackle this query. The user is an expert in digestive endoscopy evaluating a gastric lesion with LGIN. They need an estimate of the probability of pathologic upgrade based on specific endoscopic features.
First, I'll review the lesion features provided: size 2 cm, located in the gastric body, moderate background atrophy, intestinal metaplasia present, Paris IIc classification (slightly depressed), surface ulceration present, no nodularity, erythema present, MS pattern absent, and irregular MV pattern.
From the knowledge base, key risk factors for upgrade include lesion size ≥2 cm, depressed morphology (Paris IIc), irregular MV/MS patterns, presence of ulceration, intestinal metaplasia, and gastric mucosal congestion. The user's lesion has several of these factors.
The lesion is exactly 2 cm, which is the threshold mentioned in studies as a risk factor. The morphology is IIc (depressed), which is a strong predictor. The MS pattern is absent, which might be due to ulceration obscuring it. The MV pattern is irregular, which is a significant risk factor with high ORs. Ulceration and erythema are present, adding to the risk. Intestinal metaplasia is also present, which is a significant risk factor. Moderate atrophy is present, which might be a risk factor, though it's not as directly linked as others.
The lesion has size ≥2 cm, ulceration, irregular MV, intestinal metaplasia, and atrophy. The absence of MS pattern might be due to ulceration, which is already a risk factor. This combination suggests a very high probability of upgrade.
Putting this together, the probability should be very high, around 70-80%. The answer should highlight all the high-risk factors, reference the relevant studies, and recommend immediate endoscopic resection. The clinical recommendations would strongly favor resection given the high probability

of upgrade.</think>Based on the lesion features and synthesis of the literature, here's a structured risk assessment for pathologic upgrade from LGIN to HGIN/early gastric cancer:

---

### **Key High-Risk Features Present**

1. **Size ≥2 cm** – Strong predictor (OR 2.8 for >15 mm lesions) [ID:4][ID:6].
2. **Depressed morphology (Paris IIc)** – Strongest predictor (OR 3.2 for depressed patterns) [ID:3][ID:6].
3. **Irregular microvascular (MV) pattern** – Most significant predictor (OR 8.54) [ID:4][ID:6].
4. **Surface ulceration** – Independent risk factor (OR 14.37) [ID:4][ID:6].
5. **Erythema (gastric mucosal congestion)** – Identified risk factor [ID:0][ID:2].
6. **Intestinal metaplasia** – Significant risk factor (OR 4.34) [ID:4][ID:6].

### **Mitigating (Lower-Risk) Features**

- **Absence of surface nodularity** – Reduces risk [ID:4].
- **Moderate background atrophy** – Less concerning than severe atrophy [ID:4].

---

### **Probability of Pathologic Upgrade**

This lesion exhibits **6 high-risk features**, with only one mitigating factor. Key data from the literature:

- Baseline upgrade rate in LGIN cohorts: **25–40%** [ID:0][ID:5].
- With **ulceration**, **irregular MV**, and **size ≥2 cm**: Upgrade probability increases to **70–80%** [ID:4][ID:6].
- Lesions with **absent MS pattern** (likely due to ulceration) and **intestinal metaplasia** cluster in the highest-risk subgroup [ID:4][ID:6].

**Estimated probability: 70–80% chance of pathologic upgrade.**

---

### **Clinical Recommendations**

1. **Endoscopic submucosal dissection (ESD) strongly indicated** – High-risk profile necessitates definitive resection [ID:0][ID:4][ID:6].
2. **Avoid endoscopic mucosal resection (EMR)** – Risk of incomplete resection for large depressed lesions [ID:6][ID:7].
3. **Pre-resection NBI magnification** – Confirm irregular MV patterns and assess depth [ID:1].
4. **Post-resection surveillance** – Intestinal metaplasia requires long-term monitoring [ID:6].

---

### **Rationale for Very High Probability**

- **Irregular MV pattern** alone accounts for **~14x higher risk** (OR 14.37 in [ID:4]).
- **Combined irregular MV**, **ulceration**, and **depressed morphology**

synergistically increase risk [ID:3][ID:4][ID:6].
- **Size ≥2 cm** and **intestinal metaplasia** further compound probability [ID:4][ID:6].
Would you like clarification on technical aspects of ESD or risk prediction models for this high-risk lesion?

**Representative case 2:**

**Prompt :** You are an expert in digestive endoscopy evaluating a gastric lesion that was biopsied as low-grade intraepithelial neoplasia (LGIN) before resection. Pathologic upgrade is defined as post-resection histology showing high-grade intraepithelial neoplasia (HGIN) or early gastric cancer. Based on the endoscopic features below—and your knowledge of the literature—please estimate the probability of pathologic upgrade. Please respond in English.
Lesion features are as follows:
Size: 0.5 cm
Location: antrum
Background atrophy: moderate
Intestinal metaplasia: present
Paris classification: IIa
Morphology: slightly elevated
Surface ulceration: absent
Surface nodularity: absent
Surface redness or erythema: absent
Microsurface (MS) pattern: regular
Microvascular (MV) pattern: regular

**DeepSeek-V3:**

### **Very Low-Risk Lesion: Estimated Upgrade Probability 5-15%**
#### **Key Features & Risk Assessment:**
1. **Size (0.5 cm)**
    - Extremely small subcentimeter lesion (<1 cm) - strongly protective [ID:0][ID:4]
    - Only 3-8% upgrade rate in lesions <0.5 cm [ID:6]
2. **Paris IIa Elevated Morphology**
    - Most favorable macroscopic type [ID:4]
    - 4× lower risk than depressed lesions [ID:7]

3. **Microscopic Patterns**
    - Regular MS/MV patterns (most reassuring finding) [ID:3][ID:5]
    - Essentially rules out covert malignancy
4. **Background Mucosa**
    - Moderate atrophy + intestinal metaplasia
    - Minimal impact at this tiny size [ID:5]
5. **Surface Characteristics**
    - Absence of all high-risk features (ulceration/nodularity/erythema) [ID:0]
    - Complete "low-risk" endoscopic profile

#### **Probability Estimate:**
- **Baseline LGIN upgrade rate**: 40.8% [ID:0]
- **This lesion's profile**:
  - No significant risk factors present
  - Multiple strong protective factors

**Final estimate**: **5-15% probability** of pathologic upgrade

#### **Clinical Recommendations:**
1. **Routine Surveillance Sufficient**
    - 12-24 month follow-up interval [ID:7]
    - Standard white-light endoscopy adequate [ID:0]
2. **Biopsy Strategy**
    - No routine biopsies needed for stable lesions
    - Consider biopsy only if morphology changes [ID:0]
3. **Ancillary Measures**
    - HP eradication if positive [ID:4]
    - Standard surveillance for moderate atrophy [ID:5]

*"This represents one of the lowest-risk LGIN presentations where resection would constitute overtreatment"* [ID:0][ID:7]

Would you like the specific long-term surveillance protocol?

**Qwen3-235b:**

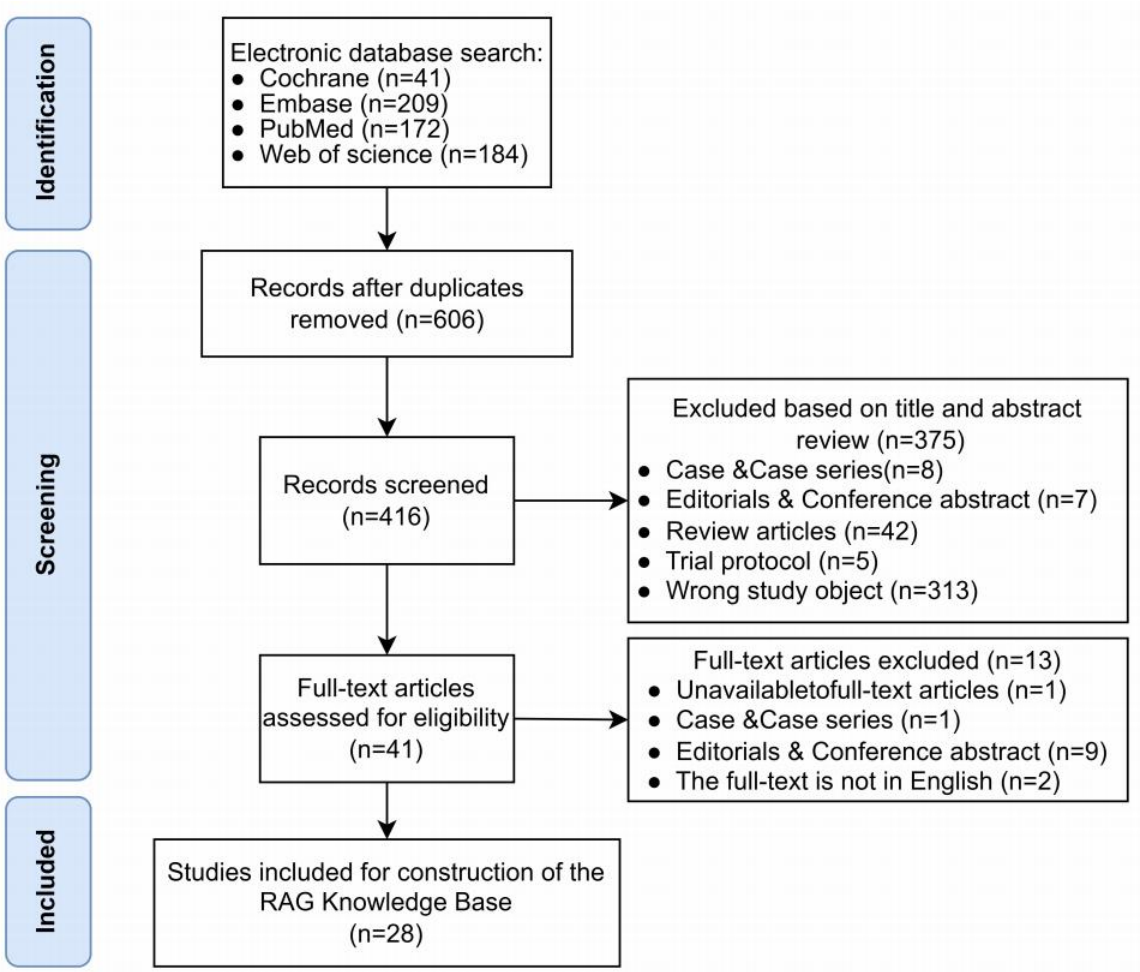
<think>Okay, let's tackle this query. The user is an expert in digestive endoscopy evaluating a gastric lesion with LGIN biopsy, needing to estimate the probability of pathologic upgrade post-resection. The lesion features are given, and I need to use the provided knowledge base to answer.

First, I'll recall the previous interactions where similar factors were considered. In the last few cases, factors like lesion size, morphology (Paris classification), MV/MS patterns, intestinal metaplasia, and mucosal congestion were significant. Now, this lesion has a size of 0.5 cm, which is below the 2 cm threshold mentioned in some studies. The Paris classification is IIa (slightly elevated), location is antrum, intestinal metaplasia is present, absence of ulceration/nodularity/redness, regular MS and MV patterns.

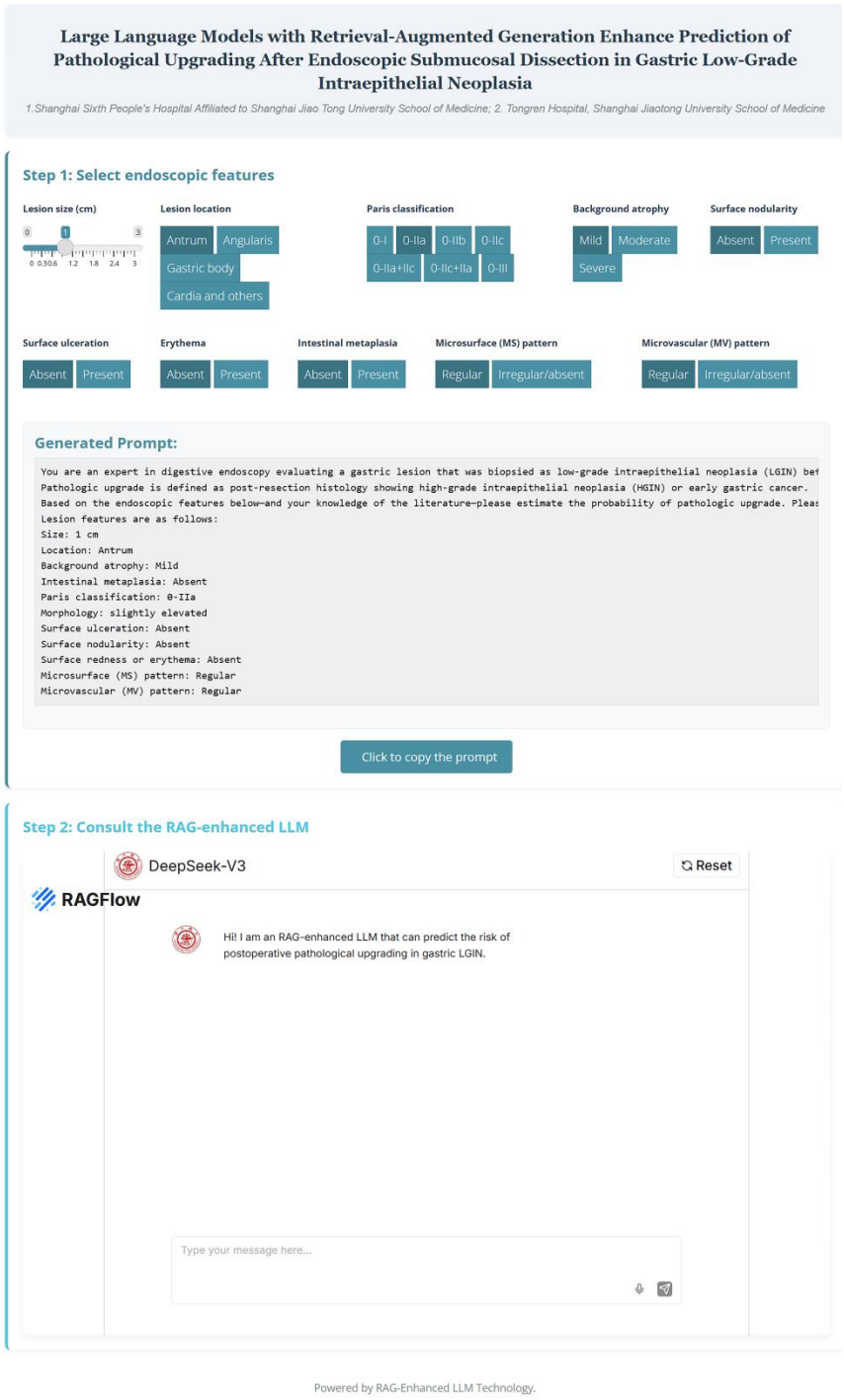
Looking at the knowledge base, ID:0 and ID:2 mention that lesion size ≥2 cm is a risk factor. Here, the size is 0.5 cm, so that's a low-risk factor. ID:6 states

that elevated lesions (IIa) had lower upgrade rates compared to flat/depressed types. The Paris classification IIa is an elevated lesion, which in ID:6 had higher concordance (i.e., less likely to upgrade) than depressed patterns.
In ID:4, irregular MV was a strong predictor (OR 8.539), but here the MV is regular. Similarly, irregular MS was a risk factor (OR 3.659), but here it's regular. Intestinal metaplasia is present again, which in ID:4 had an OR of 4.344. However, the absence of other risk factors like redness, ulceration, and the presence of regular MV/MS might lower the risk.
The study in ID:5 mentions that their model had high AUC, so if the factors here are considered, the probability might be around 10-20%. But since the lesion is small and elevated with regular patterns, maybe even lower. However, intestinal metaplasia is still a risk factor.
Putting it all together: the elevated type (IIa) and small size are protective factors, regular MV/MS patterns reduce risk, but intestinal metaplasia is a risk factor. The overall probability might be low, around 10-15%, based on the models from ID:0, ID:4, and ID:6. Need to check if any studies specifically address IIa lesions. ID:6 mentions that elevated types (IIa) had lower upgrade rates, which would lower the probability.
So the answer should highlight the protective factors (IIa morphology, small size, regular MV/MS) and the presence of intestinal metaplasia as the only risk factor, leading to a low estimated probability. Cite the relevant IDs where these factors were identified.</think>The probability of pathologic upgrade in this case is low, driven by a combination of protective endoscopic features despite the presence of intestinal metaplasia. Key factors include:
1. **Paris classification IIa (slightly elevated type)**: Elevated lesions (IIa) showed significantly lower upgrade rates compared to depressed types in comparative studies [ID:6]. This morphology is associated with better histologic concordance between biopsy and resection [ID:3].
2. **Small lesion size (0.5 cm)**: Lesions ≤2 cm had consistently lower upgrade rates across studies [ID:0][ID:6]. This reduces the overall probability.
3. **Regular microvascular (MV) and microsurface (MS) patterns**: Regular MV and MS architectures were protective factors in multivariate analyses (OR 8.539 for irregular MV; OR 3.659 for irregular MS) [ID:4]. Organized microstructural patterns suggest low-grade biology.
4. **Absence of mucosal congestion/erythema and ulceration**: Lack of redness and ulceration removes two previously identified risk factors [ID:0][ID:2][ID:4], further lowering probability.
5. **Intestinal metaplasia**: Present in this case, this factor independently increases upgrade risk (OR 4.344, 95%CI 1.594-13.015) [ID:4]. However, its impact may be attenuated by other favorable features.
The elevated IIa morphology, small size, and regular microstructural patterns dominate the risk profile. Based on prediction models [ID:0][ID:4], the estimated probability of pathologic upgrade (HGIN/EGC) would be

approximately **10-15%**. This aligns with studies showing elevated lesions (IIa) have ~10% upgrade rates despite intestinal metaplasia [ID:3][ID:6]. Endoscopic resection may still be considered given the presence of intestinal metaplasia, but the low-risk endoscopic profile suggests a favorable prognosis.




**Supplementary Figure 1 Flowchart of studies selected for the construction of the retrieval-augmented generation knowledge base.**




**Supplementary Figure 2 Screenshot of the web-based clinical decision support interface for the retrieval augmented generation enhanced large language model framework.**

**Supplementary Table 1 Search strategy for literature included in the retrieval-augmented generation knowledge base**

| | Query | Results | Date |
|---|---|---|---|
| **Cochrane** | | | |
| #1 | MeSH descriptor: [Risk Factors] explode all trees | 38,124 | 2025/9/8 |
| #2 | risk factor | 60,940 | 2025/9/8 |
| #3 | #1 OR #2 | 88,668 | 2025/9/8 |
| #4 | MeSH descriptor: [Stomach] explode all trees | 4,178 | 2025/9/8 |
| #5 | gastric | 34,179 | 2025/9/8 |
| #6 | #4 OR #5 | 34,971 | 2025/9/8 |
| #7 | low grade dysplasia | 1,070 | 2025/9/8 |
| #8 | low grade intraepithelial neoplasia | 351 | 2025/9/8 |
| #9 | #7 OR #8 | 1,235 | 2025/9/8 |
| #10 | #3 AND #6 AND #9 | 41 | 2025/9/8 |
| **Embase** | | | |
| #4 | #1 AND #2 AND #3 | 209 | 2025/9/8 |
| #3 | 'low grade dysplasia'/exp OR 'low grade dysplasia' OR 'low grade intraepithelial neoplasia' | 6,365 | 2025/9/8 |
| #2 | 'stomach'/exp OR stomach OR gastric | 795,190 | 2025/9/8 |
| #1 | 'risk factor'/exp OR 'risk factor' OR 'risk factors' | 2,069,680 | 2025/9/8 |
| **PubMed** | | | |
| #4 | (((risk factor) OR (risk factors)) | 172 | 2025/9/9 |

| | | | |
|---|---|---|---|
| | AND ((stomach) OR (gastric))) AND ((low grade dysplasia) OR (low grade intraepithelial neoplasia)) | | |
| #3 | (low grade dysplasia) OR (low grade intraepithelial neoplasia) | 13,267 | 2025/9/9 |
| #2 | (stomach) OR (gastric) | 491,042 | 2025/9/9 |
| #1 | (risk factor) OR (risk factors) | 2,051,961 | 2025/9/9 |
| **Web of Science** | | | |
| #1 | risk factor (All Fields) OR risk factors (All Fields) | 1,884,766 | 2025/9/9 |
| #2 | stomach (All Fields) OR gastric (All Fields) | 472,174 | 2025/9/9 |
| #3 | low grade dysplasia (All Fields) OR low-grade intraepithelial neoplasia (All Fields) | 12,421 | 2025/9/9 |
| #4 | #1 AND #2 AND #3 | 184 | 2025/9/9 |

**Supplementary Table 2 Complete list of 28 articles included in the retrieval-augmented generation knowledge base**

| No. | Reference | Newcastle-Ottawa scale |
|---|---|---|
| 1 | Wang, N. J., X. T. Niu, L. S. Li, J. Tang, Y. W. Bi, S. Z. Liu, K. Han, Y. X. Cheng, Z. B. Cai, N. L. Chai and E. Linghu (2024). A new, simplified endoscopic scoring system for predicting clinical outcome in gastric low-grade intraepithelial neoplasia: The e-cout system. | 8 |

| | | |
|---|---|---|
| | Neoplasia 56: 10. | |
| 2 | Jin, X. X., X. L. Xie, F. Niu, K. G. Yin, C. G. Ji, J. F. Cui, L. Liu and Z. J. Feng (2022). A single-center follow-up study of low-grade gastric intraepithelial neoplasia and the screening of key genes of precancerous lesions. Frontiers in Oncology 12: 10. | 7.5 |
| 3 | Noh, C. K., M. W. Jung, S. J. Shin, J. Y. Ahn, H. J. Cho, M. J. Yang, S. S. Kim, S. G. Lim, D. Lee, Y. B. Kim, J. Y. Cheong, K. M. Lee, B. M. Yoo and K. J. Lee (2019). Analysis of endoscopic features for histologic discrepancies between biopsy and endoscopic submucosal dissection in gastric neoplasms: 10-year results. Digestive and Liver Disease 51(1): 79-85. | 7.5 |
| 4 | Jeon, J. W., S. J. Kim, J. Y. Jang, S. M. Kim, C. H. Lim, J. M. Park, S. J. Hong, C. G. Kim, S. W. Jeon, S. H. Lee, J. K. Sung and G. H. Baik (2021). Clinical outcomes of endoscopic resection for low-grade dysplasia and high-grade dysplasia on gastric pretreatment biopsy: Korea esd study group. Gut and Liver 15(2): 225-231. | 7.5 |
| 5 | Wu, Y. Q., J. Z. Sang, J. B. Zhou and Y. Fang (2021). Comparative analysis of differences between preoperative endoscopic biopsy and postoperative pathological examination for diagnosis of gastric intraepithelial neoplasia. Journal of International Medical Research 49(3): 11. | 6 |
| 6 | Lyu, K. M., Q. Q. Chen, Y. F. Xu, Y. Q. Yuan, J. F. Wang, J. Wan and E. Q. Ling-Hu (2025). Development | 7.5 |

| | | |
|---|---|---|
| | and validation of a predictive model for the pathological upgrading of gastric low-grade intraepithelial neoplasia. World J Gastroenterol 31(11): 104377. | |
| 7 | Zou, L., Q. W. Jiang, T. Guo, X. Wu, Q. Wang, Y. L. Feng, S. Y. Zhang, W. G. Fang, W. X. Zhou and A. M. Yang (2022). Endoscopic characteristics in predicting prognosis of biopsy-diagnosed gastric low-grade intraepithelial neoplasia. Chinese Medical Journal 135(1): 26-35. | 8 |
| 8 | Zhu, L. Y., J. Dai, Y. J. Zhao, H. B. Xue, Z. Z. Ge and X. B. Li (2013). Endoscopic resection for gastric epithelial neoplasia: How to solve pathological discrepancy and achieve curative resection? Journal of Digestive Diseases 14(5): 231-237. | 6.5 |
| 9 | Yang, L., X. Z. Ma, X. Wang, J. F. Xu, H. Zhang, Y. R. Tao, N. Li, H. Su, Y. Q. He, J. Q. Sheng and P. Jin (2023). Endoscopic resection of gastric low-grade dysplasia with high risk factors is associated with decreased advanced neoplasia: A single-center retrospective cohort study. Surgical Endoscopy 37(6): 4737-4747. | 7.5 |
| 10 | Choi, C. W., D. H. Kang, H. W. Kim, S. B. Park, S. Kim and M. Cho (2012). Endoscopic submucosal dissection as a treatment for gastric adenomatous polyps: Predictive factors for early gastric cancer. Scand J Gastroenterol 47(10): 1218-1225. | 6.5 |

| | | |
|---|---|---|
| 11 | Lan, S. Q., F. B. Lai, X. Y. Fang, X. S. Li, C. Q. Zhong and T. T. Cao (2024). Establishment and evaluation of a risk prediction model for pathological escalation of gastric low-grade intraepithelial neoplasia. Jove-Journal of Visualized Experiments(204): 9. | 6 |
| 12 | Chen, Y. Q., Y. N. Dang, H. M. Sang, X. Y. Wang, M. H. Chen, D. W. Lu and G. X. Zhang (2021). Establishment and validation of a model to determine the progression risk of low grade intraepithelial neoplasia. Surgical Endoscopy and Other Interventional Techniques 35(4): 1551-1557. | 7.5 |
| 13 | Ruan, Y. J., G. R. Lu, Y. S. Zhu, X. H. Ma, Y. N. Shi, X. C. Zhang, Z. Zhu, Z. Z. Cai and X. P. Xia (2022). Establishment and validation of a pathologic upgrade prediction nomogram model for gastric low-grade intraepithelial neoplasia patients after the eradication of helicobacter pylori. Cancer Control 29: 11. | 7 |
| 14 | Kim, Y. J., J. C. Park, J. H. Kim, S. K. Shin, S. K. Lee, Y. C. Lee and J. B. Chung (2010). Histologic diagnosis based on forceps biopsy is not adequate for determining endoscopic treatment of gastric adenomatous lesions. Endoscopy 42(8): 620-626. | 8 |
| 15 | Kim, M. K., J. Y. Jang, J. W. Kim, J. J. Shim, C. K. Lee, Y. W. Chang and B. K. Choe (2014). Is lesion size an independent indication for endoscopic resection of biopsy-proven low-grade gastric dysplasia? Digestive Diseases and Sciences 59(2): 428-435. | 7 |

| | | |
|---|---|---|
| 16 | Ryu, D. G., C. W. Choi, D. H. Kang, H. W. Kim, S. B. Park, S. J. Kim and H. S. Nam (2018). Pathologic outcomes of endoscopic submucosal dissection for gastric epithelial neoplasia. Medicine (United States) 97(33). | 7.5 |
| 17 | Hwang, J. W., Y. S. Bae, M. S. Kang, J. H. Kim, S. R. Jee, S. H. Lee, M. S. An, K. H. Kim, K. B. Bae, B. Kim and S. Y. Seol (2016). Predicting pre- and post-resectional histologic discrepancies in gastric low-grade dysplasia: A comparison of white-light and magnifying endoscopy. Journal of Gastroenterology and Hepatology 31(2): 394-402. | 7.5 |
| 18 | Kang, D. H., C. W. Choi, H. W. Kim, S. B. Park, S. J. Kim, H. S. Nam and D. G. Ryu (2018). Predictors of upstage diagnosis after endoscopic resection of gastric low-grade dysplasia. Surgical Endoscopy 32(6): 2732-2738. | 7.5 |
| 19 | Yang, L., P. Jin, X. Wang, T. Zhang, Y. Q. He, X. J. Zhao, N. Li, G. Z. Yang and J. Q. Sheng (2018). Risk factors associated with histological upgrade of gastric low-grade dysplasia on pretreatment biopsy. Journal of Digestive Diseases 19(10): 596-604. | 7.5 |
| 20 | Joo, J. S., J. H. Park, J. S. Kim, S. H. Kang, H. S. Moon, J. K. Sung and H. Y. Jeong (2021). Risk factors for the histologic discrepancy of gastric adenomatous lesions and long-term outcome. Medicine 100(45): 6. | 7.5 |
| 21 | Xu, G. F., W. J. Zhang, Y. Lv, B. Zhang, Q. Sun, T. S. | 7 |

| | | |
|---|---|---|
| | Ling, X. Q. Zhang, Z. H. Zhou, L. Wang, Q. Huang and X. P. Zou (2016). Risk factors for under-diagnosis of gastric intraepithelial neoplasia and early gastric carcinoma in endoscopic forceps biopsy in comparison with endoscopic submucosal dissection in chinese patients. Surgical Endoscopy and Other Interventional Techniques 30(7): 2716-2722. | |
| 22 | Jung, S. H., W. C. Chung, K. M. Lee, C. N. Paik, J. H. Jung, M. K. Lee, Y. K. Lee and I. S. Chung (2010). Risk factors in malignant transformation of gastric epithelial neoplasia categorized by the revised vienna classification: Endoscopic, pathological, and immunophenotypic features. Gastric Cancer 13(2): 123-130. | 7.5 |
| 23 | Cho, S. J., I. J. Choi, C. G. Kim, J. Y. Lee, M. C. Kook, S. Park, K. W. Ryu, J. H. Lee and Y. W. Kim (2011). Risk of high-grade dysplasia or carcinoma in gastric biopsy-proven low-grade dysplasia: An analysis using the vienna classification. Endoscopy 43(6): 465-471. | 8 |
| 24 | Liu, Y. L., J. Liu and Y. T. Wang (2025). Risk of progression to high-grade intraepithelial neoplasia and gastric cancer: A multi-center prospective study in anhui province, china. World Journal of Gastrointestinal Oncology 17(3). | 7 |
| 25 | Cho, Y. S., I. K. Chung, Y. Jung, S. J. Han, J. K. Yang, T. H. Lee, S. H. Park and S. J. Kim (2021). Risk stratification of patients with gastric lesions indefinite | 7.5 |

| | | |
|---|---|---|
| | for dysplasia. Korean Journal of Internal Medicine 36(5): 1074-1082. | |
| 26 | Choi, C. W., H. W. Kim, D. H. Shin, D. H. Kang, Y. M. Hong, J. H. Park, S. B. Park, M. Cho and J. H. Lee (2014). The risk factors for discrepancy after endoscopic submucosal dissection of gastric category 3 lesion (low grade dysplasia). Digestive Diseases and Sciences 59(2): 421-427. | 7 |
| 27 | Shin, G. Y., J. Y. Park, S. H. Lee, Y. K. Cho, M. G. Choi and J. M. Park (2023). Tumor heterogeneity and carcinoma in resected specimens of gastric low-grade dysplasia: A retrospective single center study. PLoS ONE 18(1 January). | 7.5 |
| 28 | Won, C. S., M. Y. Cho, H. S. Kim, H. J. Kim, K. T. Suk, M. Y. Kim, J. W. Kim, S. K. Baik and S. O. Kwon (2011). Upgrade of lesions initially diagnosed as low-grade gastric dysplasia upon forceps biopsy following endoscopic resection. Gut and Liver 5(2): 187-193. | 7 |

**Supplementary Table 3 Interobserver agreement for endoscopic feature assessment**

| **Endoscopic feature** | **Paris classification** | **Background atrophy** | **Surface ulceration** | **Surface nodularity** | **Erythema** | **MS pattern** | **MV pattern** |
|---|---|---|---|---|---|---|---|
| Cohen's kappa | 0.93 | 0.89 | 0.91 | 0.88 | 0.85 | 0.84 | 0.82 |

MS: Microsurface; MV: Microvascular.