

# ASUKAI89 at NTCIR 18 RadNLP : Lung Cancer Staging Automatic Classification System Utilizing Large Language Models and Meta-Prompting

Yoshifumi Okura<sup>1)</sup> Yuki Kataoka<sup>1) 2)</sup>

1) Kyoto Min-iren Asukai Hospital  
2) Scientific Research Works Peer Support Group(SRWS-PSG)

Using large language models and meta prompts, RadNLP automatically extracts TNM classification from radiology reports with 70% joint accuracy, surpassing rule-based baselines. Further data augmentation will bring it closer to clinical deployment.

## BACK GROUND

Recent LLM approaches show promise in:

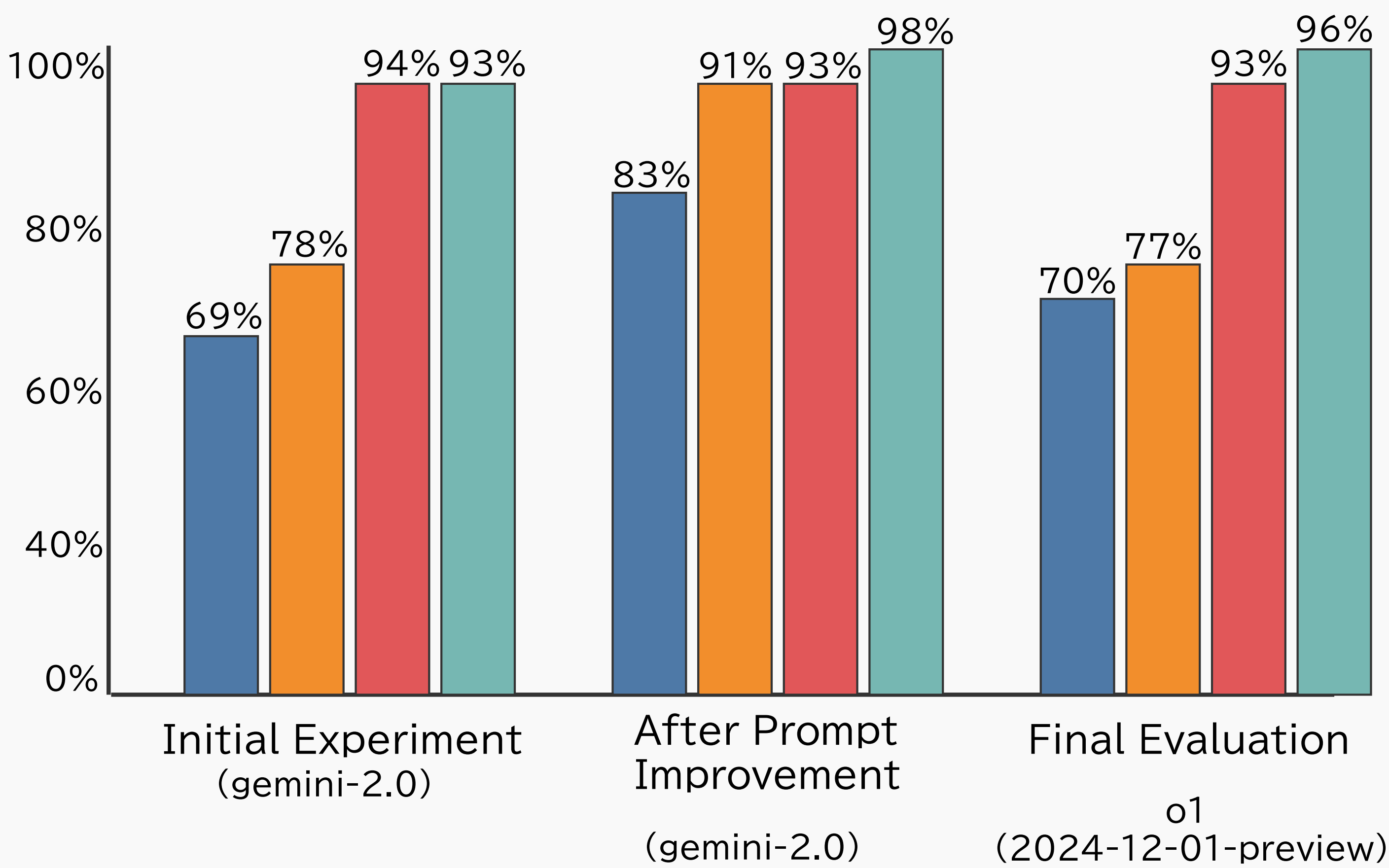
- Yamagishi et al. developed zero-shot information extraction for Japanese radiological reports
- Chia et al. demonstrated effective extraction of pTNM classification from pathology reports using proper prompt design

This study improves automated TNM classification extraction using prompt engineering and meta-prompting

## Meta-Prompting Example Table

Aspect	Initial Prompt	Meta-Prompted Improvement
Unit Standardization	Used "cm" units (e.g., "≦ 3cm")	Converted all measurements to "mm" for consistency
T-factor Decision Process	No explicit decision steps	Added structured verification pathway for T classification
Error Prevention Logic	No error checking mechanism	Implemented mandatory final verification checklist

## RESULTS



- Meta-prompting improved accuracy: +14.8% Joint, +12.9% T accuracy
- N factor and M factor achieved high accuracy
- T factor accuracy is lower likely due to multiple tumor parameters requiring interpretation and ambiguous descriptions in radiological reports

## FUTURE WORK

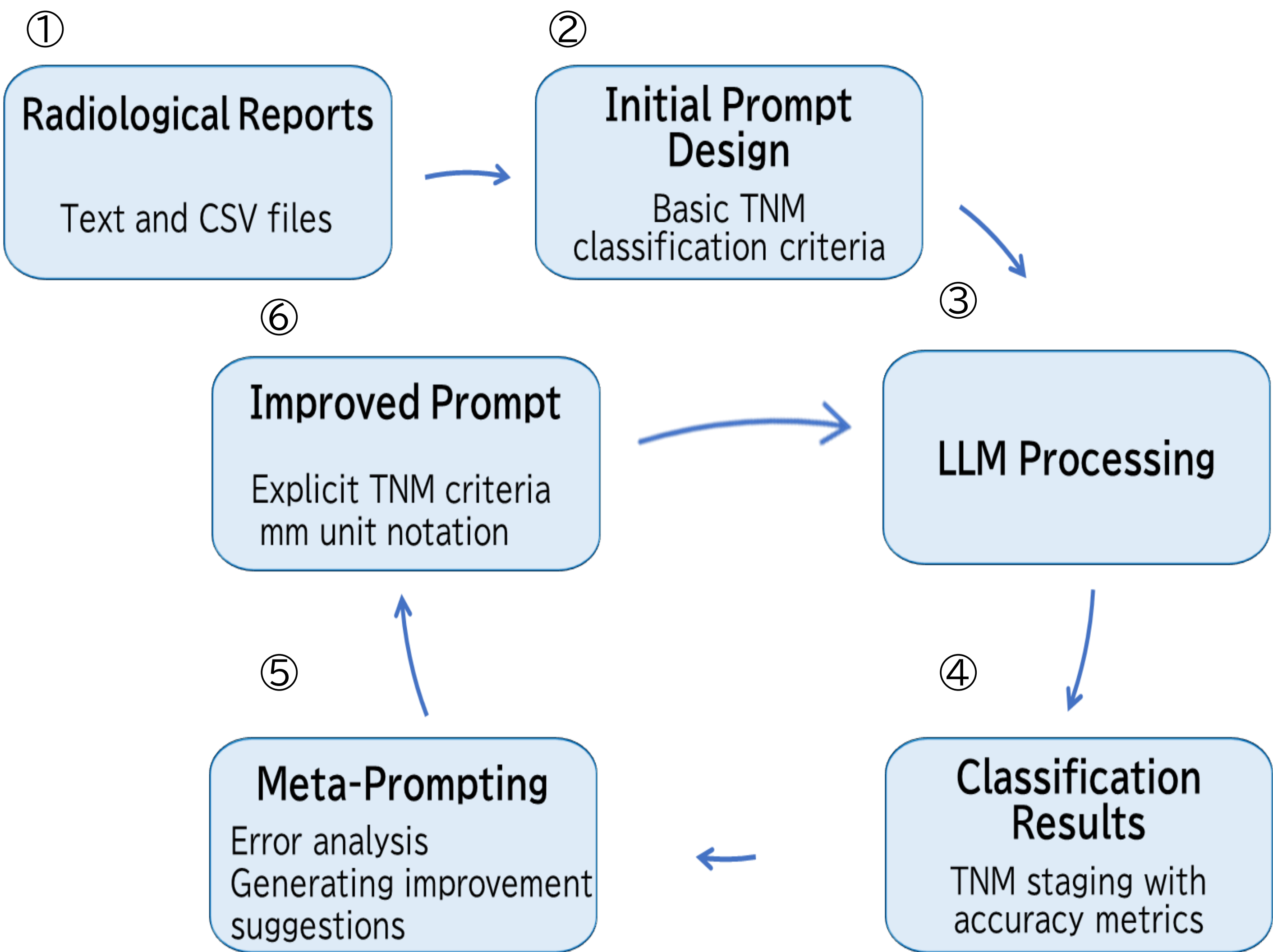
- Validation with larger datasets
- Development of multilingual capabilities
- Creating integrated approaches for T-N-M factor interdependencies

## FURTHER READING

Scan this QR code to access our GitHub repository containing the complete source code, detailed prompt examples, and implementation details of our RadNLP system.



## METHODS



- Addition of explicit TNM classification criteria
  - Clearly stated TNM classification criteria at the beginning of the prompt
- Standardization of unit notation
  - Improved consistency with measurement expressions in reports
- Error analysis and iterative improvement
  - Systematic analysis of misclassifications in training data
  - Generation of improvement proposals using gemini-2.0 model
  - Final inference with o1(2024-12-17) model